Understanding Innovation

Christoph Meinel
Larry Leifer *Editors*

Design Thinking Research

Looking Further: Design Thinking Beyond Solution-Fixation

Springer

Understanding Innovation

Series editors
Christoph Meinel
Potsdam, Germany

Larry Leifer
Stanford, USA

More information about this series at http://www.springer.com/series/8802

Christoph Meinel • Larry Leifer
Editors

Design Thinking Research

Looking Further: Design Thinking Beyond Solution-Fixation

Springer

Editors
Christoph Meinel
Hasso Plattner Institute for Digital Engineering
University of Potsdam
Potsdam, Germany

Larry Leifer
Stanford University
Stanford, CA, USA

ISSN 2197-5752 ISSN 2197-5760 (electronic)
Understanding Innovation
ISBN 978-3-319-97081-3 ISBN 978-3-319-97082-0 (eBook)
https://doi.org/10.1007/978-3-319-97082-0

Library of Congress Control Number: 2018957128

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Complexity grows ubiquitously in our world, and the resultant challenges are impossible to ignore. New developments in digital technologies often suggest and frequently demand making changes—sometimes radical—to how products, tools, and services could or should function, how we collaborate and interact with customers and partners, and how we organize processes and business. In order to move forward effectively, it is necessary for us to think differently, departing from habitual and traditional ways of thinking. Design thinking has been adopted and practiced by individuals and organizations in increasing numbers as a powerful framework to foster innovation in products, services, and operations, and more recently on strategy and creating cultures of innovation.

As interest and experience in design thinking have grown, as well as its application to new and diverse challenges, there is an increasing need to deepen our understanding of how and why design thinking works: what are the factors that make it more successful than other approaches, and what are the reasons it can fail? These are among the driving questions that fuel my continuous interest and support for meticulous and conscientious research in the field through the Hasso Plattner Design Thinking Research Program, a research initiative conducted jointly by the Hasso Plattner Institute of Design at Stanford University in California and the Hasso Plattner Institute (HPI) for Digital Engineering in Potsdam, Germany.

Over the course of nearly ten years of groundbreaking research, our understanding of design thinking has broadened and deepened in an unprecedented manner. Since the implementation of the Design Thinking Research Program in 2008, researchers on two continents have conducted over 100 studies examining, detailing, and making sense of design thinking in its many forms. We now have an estimable body of new knowledge concerning team dynamics, the characteristics and mechanisms of effective design thinking tools, and the application of design thinking in organizational contexts. New tools and methods based on the knowledge and insights created by the program are laying a foundation for informed, empirically based practice that yields greater impact. It is incumbent upon the research community to channel the new knowledge we have created into improved content for teaching and learning design thinking in all HPI communities.

New discoveries and insights, newly developed approaches, and tools in design thinking should be available to all who seek to advance, drive, and support innovation and innovative culture for organizations or for individuals making social and cultural change. In light of this, it is essential for new design thinking knowledge and content—studies, workshops, frameworks, exercises, and tools—to be available and accessible to students and professionals, to experts, and to those just starting out.

Last year the tenth anniversary of the founding of the School of Design Thinking at Hasso Plattner Institute in Potsdam was celebrated at the second d.confestival. Scholars of the Design Thinking Research Program complemented the festivities by running an enthusiastically attended "Design Thinking Research Lab" in the event space. Here, visitors learned not only about new contributions to and insights into design thinking, but also about how research is conducted and why it is a force of relevant change in our communities.

At this moment, you hold in your hands another way of bringing our findings to innovators everywhere: the ninth volume of our series on design thinking research presents the comprehensive collection of research studies carried out by scholars of both the Hasso Plattner Institute in Potsdam and at Stanford University. In addition to providing the findings of recent projects, this volume presents a thought-provoking historical perspective on design thinking as a creative practice. Now in its second funding period, the Design Thinking Research Program has cultivated a diverse and growing community that gives place to rich exchanges between current doctoral candidates, alumni, researchers, and practitioners from a myriad of disciplines. These exchanges bring new perspectives, depth, and lasting value to not only the program and to its researchers, but to design thinking itself.

The synthesis of multiple perspectives—which results from teams of people from diverse disciplines, from research and from practice, collaborating on the same, often wicked challenge—constitutes one of the fundamental benefits of design thinking. In this enterprise of design thinking research, we encourage you to reach out and invite curious innovators of all kinds to work together, to experiment and learn, and to focus on broadening and deepening our practice and understanding of design thinking and the impact it can bring to benefit the world.

Palo Alto, CA Hasso Plattner
Winter 2017/2018

Contents

Looking Further: Design Thinking Beyond Solution-Fixation

Larry Leifer and Christoph Meinel

Abstract In the pursuit of breakthrough innovation, it is necessary to recognize that most, if not all, respective disciplines exhibit tendencies for "solution-fixation." When problems appear, the historical value of designers and managers has been associated with an ability to swiftly recognize, properly mobilize, and solve them. The Hasso Plattner Design Thinking Research Program (HPDTRP) redirects analytical and generative attention elsewhere: to the development and nurturing of a problem-oriented mindset.

1 Fixations and Orientations

In the pursuit of breakthrough innovation, it is necessary to recognize that most, if not all, respective disciplines exhibit tendencies for "solution-fixation." When problems appear, the historical value of designers and managers has been associated with an ability to swiftly recognize, properly mobilize, and solve them. Such assured action fulfills the general expectations of the world at large, and this prevailing behavior is further nurtured by both urgency found on the front lines of design practice and increasing demand for applied research.

The Hasso Plattner Design Thinking Research Program (HPDTRP) redirects analytical and generative attention elsewhere, while reappropriating the prevailing mindset when necessary. We find greater value in the development and nurturing of a *problem-oriented* mindset to best quell prevailing *solution-fixation* practices that continue to drive some disciplinary specific practices today. Through our engagement with found problems, any natural tendency *to solve* is first suspended to better interrogate and reframe problems for end-user gain. This temporary suspension of

L. Leifer (✉)
Stanford Center for Design Research, Stanford University, Stanford, CA, USA
e-mail: leifer@cdr.stanford.edu

C. Meinel
Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany
e-mail: meinel@hpi.de

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_1

solution generation better allows for the development of a sustained user-centered mindset.

This problem-orientation strategy is now evident across 10 years of HPDTRP projects, and lends itself as a viable tactic for all designers. While both design thinking and design thinking research is of increasing interest to many people, we continue to uncover new features and phenomena that emerge from the user-centered mindset and its application in particular context-specific design operations. User-centrism is a prerequisite for our design action, and proper problem-orientation is a requirement for generating any meaningful and worthwhile design solution.

1.1 Fixation

In his 2007 title *Designerly Ways of Knowing*, Nigel Cross characterizes *fixation* as when designers "may be too ready to re-use features of known existing designs, rather than to explore the problem and generate new design features."[1] Typically, designers are designers of something. They own and are identified with that thing, or style of thing. We prompt the designer within each of us to pause and question the problem as stated. Does this familiar problem really deserve to be re-solved? The posing of this question stems from observations in both professional and academic environments. In most cases, the constraints that define problem spaces, either shaped internally or impacted by forces extrinsic to the problem itself, prompt a quickness in the execution of design action perceived to be optimal, yet rarely proves to be. The quickness introduced by *solution-fixation* is most unhelpful, as it precludes the questioning or interrogation of problems stated by others, and prematurely freezes the problem space before fully forming.

Such fixation is further characterized by a designer's demonstrated attachment to a solution over time. While some liberties are taken with problem constraints and expectations for outcomes are adjusted, designers demonstrate attachment "to their principal solution concept for as long as possible, even when detailed development of the scheme throws up unexpected difficulties and shortcomings in the solution concept."[2] Typically, the designer in each of us tends to privilege the first solution concept that presents itself. The expenditures of time, money and effort give pause and trigger a rethinking of the issues. These expended costs may be entirely justified if breakthrough innovation is needed and a new business model warranted.

The most unfortunate shortchanging of a robust design process however is the negative effect of solution-fixation on the divergent exploration of creative possibility in design. When design activity is fixated on a particular solution concept, then

[1]Cross, Nigel. Designerly Ways of Knowing (Board of International Research in Design). Birkhauser: Basel, 2007. p104.

[2]Cross, Nigel. Designerly Ways of Knowing (Board of International Research in Design). Birkhauser: Basel, 2007. p105.

the divergent generation of design options and alternatives will seem "to be in conflict with a more 'principled' approach to design and even conflicts with the idea that it is the exploration of solution concepts that assists the designer's problem understanding."[3] Typically, we are time challenged and efficiency inclined to privilege first ideas, as the creation of viable options and alternatives certainly smells inefficient. Research in our community however yields dramatic evidence that designers, challenged to deliver three to five alternative prototypes, deliver solutions that work 3× to 5× better than when asked to deliver a single "solution/prototype."[4] In this light, solution-fixation behavior, of which we are all guilty of exhibiting at times, betrays the primary value proposition that designers offer.

For those contexts in which these effects are undesirable, one will want to actively protect against exhibiting such known practices, and guard against their invisible presence. As the liabilities of *solution-fixation* are numerous, our challenge then is to be mindful of its causes. In turn, these causes are invariably tied to how problems themselves are first identified and managed. Fortunately, design researcher Kees Dorst has identified and named five causal syndromes of conventional problem-solving practices.[5] These are:

- *Lone Warrior*: Some one entity owns the problem-solution space.
- *Freeze the World*: Stop the world, prevent change, utilize static thinking.
- *Self-Made Box*: Solve the problem with solutions from our past.
- *Rational High Ground*: Believe in, and assert our own rationality.
- *Identification*: The problem and its solution are identified through organizational autopoiesis.

Proceed with extreme caution when you detect these in your respective environment or culture. For instance, while problems requiring software solutions can be reasonably updated in a short amount of time, problems of the built environment require solutions with physical configurations lasting years, decades or centuries.

1.2 Orientation

Developing an awareness of *solution-fixation,* and recognizing the evidence it yields, enables a safeguarding against it. This awareness is especially helpful during design scenarios in which we self-detect. By developing greater self-awareness, designers can adjust and compensate for any solution-fixated design behavior in response to

[3]Cross, Nigel. Designerly Ways of Knowing (Board of International Research in Design). Birkhauser: Basel, 2007. p106.

[4]Klemmer, Scott. *CHI 2011: Conference on Human Factors in Computing Systems*. May 7–12, 2011, Vancouver, BC, Canada. Copyright 2011 ACM 978-1-4503-0267-8/11/05.

[5]Dorst, Kees. Frame Innovation: Create New Thinking By Design. MIT Press: Cambridge, 2015. p13–17.

the type and volume of design materials they generate.[6] Because of the inherent quickness with which fixation occurs, awareness is a prerequisite for *problem-orientation*.

The pursuit of any problem-oriented design action however requires the development of a human-centered mindset. In turn, problem orientation is then an accurate indicator of whether the mindset guiding action is in fact human-centered, or instead is rooted within any number of other centrisms. Examples include systems-centrism, type-centrism, or ego-centrism.

There are multiple strategies for pivoting towards problem-orientation. These include:

- *Identify meaningful problems*. Does the problem identified have consequence, or is it inconsequential? Is it significant and relevant to stakeholders beyond the designers? Can the problem be reframed in multiple ways and still be of relevance to multiple stakeholding groups?
- *Phrase effective questions*. Do the questions posed establish a larger context for the problem? Do the questions reveal details of the problem either invisible or previously unknown? Are the questions appropriately divergent and convergent for framing the design problem?
- *Develop "comprehensive propensities" for design action*. Is it possible to know the entirety of a problem within a single discipline, or is the scale of the problem large enough in which no single stakeholder can perceive it in its natural entirety? Does this problem of perception change as soon as one views it in a disciplinarily-agnostic way?

1.3 The Applicability of Design Research

How might we make these distinctions between fixation and orientation be actionable on a day-to-day, session-to-session basis within the enterprise? Can an exceptional culture of extreme problem-orientation positively affect the prevailing culture of solution-fixation?

Can we summarize the challenge as the distinction between freshly asking "what is the problem really?" versus pushing forward with familiar or pre-conceived solutions to recognizable but ill-identified problems?

Is increasing emphasis to *solve* introducing rewards for systematized statement making and systematized design engagement, thereby generating new problems exacerbating the original one?

Can we pivot skillfully from solution-fixation to problem-orientation while remaining civil?

[6] Ford, Chris. Email correspondence with author, March 30, 2018.

Design Research enables our ability to address these questions and others like them, with new metrics and a heightened awareness of the un-intended bias at the core pursuit of breakthrough innovation in business, government, and academia. It promises to also hold great value while navigating new creative possibilities by prioritizing human-centered problem formulation. For design operations to find greater resonant meaning and impact, they must feature human users at the center of (increasingly digitally supported) endeavors, including the creation of new products, services, processes, and systems.

2 The HPI-Stanford Design Thinking Research Program

Design thinking as a user-centric innovation method has become widespread during recent years in practice, education, and academia. A growing number of people and organizations have experienced its innovative power. At the same time the demand to understand this method has also increased. In 2008 the joint HPI Stanford Design Thinking Research Program was established, funded by the Hasso Plattner Foundation. Within this program, scientists, designers, and humanists from the Hasso Plattner Institute for Digital Engineering (HPI) in Potsdam, Germany, and from Stanford University, USA, gain a deeper research-enabled understanding of the underlying principles of design thinking and, consequently, how and why this innovation paradigm succeeds or fails.

2.1 Program Vision and Goals

Multidisciplinary research teams from HPI and Stanford with backgrounds in disciplines such as engineering, design, humanities, or social sciences investigate innovation and design thinking in a holistic way. These areas of investigation center on technical, economic, and human factors. Applying rigorous academic methods, researchers examine how the design thinking paradigm can be improved and further developed.

The program pursues the goal to advance design thinking theory and knowledge within the research community and ultimately improve design practice and education by funding original research to support design activities. It seeks to yield deep insights into the nature of human needs and the protocols that design thinking researchers might apply to achieve "insights" versus "data." Beyond a descriptive understanding of the subject matter, this program assists the development of metrics that allow an assessment and prediction of team performance to facilitate real-time management of how teams work. Researchers study the complex interaction between members of multi-disciplinary teams, with special regard to the necessity of creative collaboration across spatial, temporal, and cultural boundaries. They design, develop, and evaluate innovative tools and methods that support teams in their creative work. The research projects address questions of why structures of successful design thinking teams differ substantially from traditional corporate structures and how design thinking methods mesh with traditional engineering and management approaches.

Researchers are especially encouraged to develop ambitious, long-term explorative projects that integrate technical, economical, as well as psychological points of view using design thinking tools and methods. Field studies in real business environments are useful to assess the impact of design thinking in organizations and if any transformations of the approach may be warranted.

Special interest is found in the following questions:

– What are people really thinking and doing when they are engaged in creative design innovation?
– How can new frameworks, tools, systems, and methods augment, capture, and reuse successful practices?
– What is the impact of design thinking on human, business, and technology performance?
– How do the tools, systems, and methods work together to create the right innovation at the right time? How do they fail?

Over 129 research projects since 2008, our understanding of this field has advanced with the authoring of new tools and yielding of new insights. This *Design Thinking Research* series shares scholarly insights with a public audience, whether in a multi-national corporation or a garage-based start-up.

2.2 Road Map Through This Book

With funding for the HPI-Stanford Design Thinking Research Program renewed, we re-dedicate ourselves to make the outcomes of our work more broadly known.

Now in its ninth program year, researchers from HPI and Stanford University have conducted a wide range of research projects on design thinking. This annual publication is a compilation of their findings, sharing outcomes arranged into three parts that illustrate a comprehensive approach to design thinking research. At the beginning of this publication you will again find a historic perspective. Building upon the creative thinking theories of Stanford educator John E. Arnold first investigated last year, "Theoretical Foundations of Design Thinking: Robert H. McKim's Need-Based Design Theory" by Julia P. A. von Thienen, William J. Clancey, and Christoph Meinel focuses on the need-based design theory propounded by one of Arnold's successors. His theory, first published in 1959, advanced human-centered design conceptions by providing an elaborate account of human needs, by clarifying the role of designs and designers in the process of culture development, and by providing guidelines to assess, or actively increase, design value. According to McKim, the ultimate purpose of design is to promote the well-being of people by helping to gratify their basic needs. As his overall design framework is broadly scoped, it can also serve as a frame of reference to analyze and compare different present-day approaches to design, such as innovation-focused design thinking and usability-focused studies in Human-Computer Interaction.

The section "Understanding Success Factors of Design Thinking" shares a deeper understanding of the three vital factors for successful design thinking (DT): team collaboration, the DT-process and an innovation inviting space. To deepen our understanding of collaboration in teams, the first chapter examines the creative interaction of DT teams. It introduces the most important socio-emotional factors for creative teamwork. Further, it examines how the emotional dynamics of the team is shaped by the alternating diverging and converging phases of the design thinking process. Describing DT as a practice that can vary greatly depending on context is at the center of the second chapter. It outlines a series of measures developed to highlight both the different aspects of design thinking and how these aspects can vary, thereby enabling a more detailed assessment of what is necessary to successfully apply DT. The last chapter explores space as means to effectively foster creativity and innovation in organizations, using the concept of spatial interventions. The approaches and findings support strategists as well as practitioners and contribute to a deeper understanding of how to use space as a strategic tool.

New digital technologies allow for vital changes in the way our organizations and our economy work. The second section titled "Exploring the Digital Potential—Teaching, Research and Organizational Approaches" investigates the scaling potential of design thinking in teaching and research through digital means. Moreover, it reflects on redesign at the level of social organization fostering the requisite sociological and psychological transformations of digital shifts efficiently. The first chapter describes the iteration process and first results of the public DT-MOOC, which ran on *openHPI* with more than 3000 learners participating actively. The scaling potential of the work conducted in university laboratories is the focus of the second chapter, presenting Crowd Research, a crowdsourcing technique that coordinates open-ended research through an iterative cycle composed of open contribution, synchronous collaboration, and peer assessment. Over the course of 2 years and three projects, this crowd research has produced articles at top-tier Computer Science venues, and participants have joined leading institutions. The final chapter aims to develop solutions for accelerating the innovation rate of the new digital economy by redesigning the social organization. The shift from the hierarchical, clockwork, and command and control organizations of the industrial age is immense compared against the variety of prevalent structures designed to meet the demands of frequent and rapid change. This shift is explored with a view to the alternative redesign of social organizations and the means to accomplish the requisite sociological and psychological transformations efficiently.

The third and final section, "Design Thinking in Practice," focusses on the application of design thinking in diverse organizational environments facing different challenges. We see its fruitful application in the health care sector, in software development and in novice design. The first chapter focuses on the ongoing development process of the digital documentation system Tele-Board MED (designed for medical encounters), introducing a software that enables recording data with the help of handwritten and spoken notes. The system allows for medical personnel to better connect with the patient. Personnel can fulfill their obligations, such as writing medical reports or documenting treatment, without the patient interaction disturbed by the distraction of a keyboard or desktop monitor. In the second chapter, researchers present the application of DT in health information technology systems engineering, to improve communication between patients and providers. The team aimed to transform pain management from a unidimensional construct measured on traditional pain scales to a social transaction between patients and caregivers. They tested their approach with patients and health care professionals at Stanford Health Care. In the third chapter, the research team explores the potential that lies in combining parts of the DT process with other approaches. It explains the development of InnoDev, which is an approach that combines Design Thinking, Lean Startup and Scrum to create an agile software development process that delivers the innovative customer-oriented products and services required by competitive companies. The fourth chapter proposes a new perspective on the environments used in software development, called "exploratory software design environments." It describes the properties of such an environment and illustrates the perspective with existing related tools and environments. In the final chapter, a user-centered tool for recognizing good design proves that with the right set-up, novice designers have the ability to recognize good design. This research informs better design tools, and the creation of a meaningful design scale that can be created from more easily obtainable novice comparisons.

2.3 Part I: Understanding Success Factors of Design Thinking

The creative interaction of a team is where most of the innovation work in organizations happens. Yet the creative team is an exceptionally messy place with regard to socio-emotional interactions. Working creatively means constantly navigating and negotiating both uncertainty and ambiguity, which apart from constant procedural adaptations, both evokes and needs adequate responses on the socio-emotional level. In "Emotions along the Design Thinking Process" Axel Menning, Benedikt Ewald, Claudia Nicolai, and Uli Weinberg identify the most important socioemotional factors for creative teamwork and how the emotional dynamics of the team are shaped by the different phases of the design thinking process. To this end automated text analysis of design thinking team meetings is reviewed as a method to unobtrusively track emotional dynamics throughout the design process.

Design thinking is often misconceived as a competency that a person or organization either has or does not have. This is problematic because that perspective can lead to an incorrect assumption that design thinking is uniformly applied at the same level by anyone who knows it. Adam Royalty, Helen Chen, Bernard Roth, and Sheri Sheppard focus on "Measuring Design Thinking Practice in Context", describing design thinking as a practice that can range greatly depending on context. The chapter outlines a series of measures developed to highlight both different aspects of design thinking and how those aspects can vary. These measures provide a more detailed assessment of what is necessary to successfully apply design thinking.

In "Making use of Innovation Spaces—Towards a Framework of Strategizing Spatial Interventions" Marie Klooker, Martin Schwemmle, Claudia Nicolai, and Ulrich Weinberg explore the use of workshop space as a means of effectively fostering creativity and innovation in organizations. They shed light on crucial aspects of how innovation spaces can be used as a 'silent coach', taking into consideration the conceptual interplay of the strategic discourse, theoretical accounts of coaching practice and the process of using innovation spaces. Further, the authors introduce a framework for spatial interventions that helps to structure and analyze the use of space during a workshop. Finally, based on findings from a case study that was conducted in the newly created innovation space of a large company, this theoretical framework gets applied and expanded. The approaches and findings of this chapter support both strategists and practitioners while contributing to a deeper understanding of how to make use of space as a strategic tool.

2.4 *Part II: Exploring the Digital Potential: Teaching, Research and Organizational Approaches*

How can design thinking be taught in a Massive Open Online Course (MOOC)? In "An Iterative Approach to Online Course Design: Improving a Design Research MOOC" Karen von Schmieden, Lena Mayer, Mana Taheri, and Christoph Meinel are taking an iterative approach at designing an online course about the design research phase and built a prototype MOOC for testing. They applied three measurement tools (Course Evaluation Survey, Skill Confidence Rating, and Qualitative Interviews) and categorized the collected feedback. This process resulted in fifty-seven iterative tasks, which were implemented in the public version of the MOOC. In this chapter, they describe the iteration process and first results from the public MOOC.

Research experiences in higher education are the result of a competitive process. Providing open access to research experiences would enable global upward mobility and increased diversity in the scientific workforce. How can we coordinate a crowd of diverse volunteers on open-ended research? How could a PI have enough visibility into each person's contributions to recommend them for further study? In "Crowd Research: Open and Scalable University Laboratories" Rajan Vaish, Snehalkumar (Neil) S. Gaikwad, Geza Kovacs, Andreas Veit, Ranjay Krishna, Imanol Arrieta Ibarra, Camelia Simoiu, Michael Wilber, Serge Belongie, Sharad C. Goel, James Davis, and Michael S. Bernstein present Crowd Research, a crowdsourcing technique that coordinates open-ended research through an iterative cycle of open contribution, synchronous collaboration, and peer assessment. To aid upward mobility and recognize contributions in publications, a decentralized credit system is introduced: Participants allocate credits to each other, which a graph centrality algorithm translates into a collectively-created author order.

The third chapter, "Redesigning Social Organization for Accelerated Innovation in the new Digital Economy: A Design Thinking Perspective," by Ade Mabogunje, Neeraj Sonalkar, and Larry Leifer explores new ways of social organization suited for the overall digitally agile mode of our lives. The variety of emerging organizational structures compounded by a need for quick-change structures necessitates a biological metaphor for the organization as an organism that can fold, unfold, and refold to rapidly adapting volatile environments. This radical shift from the hierarchical, clockwork, command and control organizations of the industrial age, explores a view to showing alternative redesign of social organizations.

2.5 Part III: Design Thinking in Practice

Medical personnel's everyday work has become more and more characterized by administrative tasks, such as writing medical reports or documenting a patient's treatment. The HPI research team aims to ease data entry in doctor-patient encounters. In "New Ways of Data Entry in (Health-Care) Doctor-Patient Encounters", Matthias Wenzel, Anja Perlich, Julia von Thienen and Christoph Meinel present a software tool, Tele-Board MED, that allows recording data with the help of handwritten and spoken notes that are transformed automatically to a textual format via handwriting and speech recognition. This software is a lightweight web application that runs in a web browser. It can be used on a multitude of hardware, especially mobile devices such as tablet computers or smartphones. In an initial user test, the digital techniques were rated more suitable than a traditional pen and paper approach that entails follow-up content digitization.

Nicholas Berte, Lauren Aquino Shluzas, Bardia Beigi, Moses Albaniel, Martin S. Angst, and David Pickham explore the role of design thinking to improve pain management for patients and providers. Specifically, using a design thinking approach, they aimed to transform pain management from a unidimensional construct measured on traditional pain scales to a social transaction between patients and caregivers, through recognizing the behavioral, psychosocial, and environmental aspects of pain. In "Design Thinking Pain Management: Tools to improve human-centered communication between patients and providers", a two-phase study first develops a pain assessment intervention in the form of a novel Android-based pain management application. Initial findings demonstrate that patients and nurses were able to communicate pain needs through the use of the novel application.

The debate of how to integrate Design Thinking and Lean Startup into the agile process has been addressed in previous scholarly literature. Researchers argue that Design Thinking can contribute to software development by offering support on how to understand user needs in order to derive solution and product options, whereas Lean Startup helps learn about business and scaling strategies. Based on that, Franziska Dobrigkeit, Danielly de Paula, and Matthias Uflacker developed InnoDev, which is an approach that combines Design Thinking, Lean Startup and Scrum to create an agile software development process that can deliver the innovative customer-oriented products and services required by competitive companies. In "InnoDev—a Software Development Methodology integrating Design Thinking, Scrum and Lean Startup" the researchers describe InnoDev in detail. Their findings provide complementary perspectives regarding software development strategies, roles and techniques. This study advances the knowledge of Design Thinking and software development by providing a detailed description of a tool that combines best practices for creating more innovative software-products. The results of this investigation can help managers to evaluate their software development process to improve its effectiveness and create effective user driven solutions.

How to create a new software system can be a wicked problem. Consequently, it is important for such projects to have a collaborating team of experts from multiple disciplines. While agile development processes foster such a collaboration on the social level, the tools used by individual experts still prevent team members from seeing the overall result of their collective modifications on the resulting system. Roles in the process, such as content designers and user experience designers, only see the impact of their changes on their artifacts. Based on the concept of exploratory programming environments, Patrick Rein, Marcel Taeumel, and Robert Hirschfeld propose a new perspective on the environments used in software development, called exploratory software design environments. In "Towards Exploratory Software Design Environments for the Multi-disciplinary Team" the researchers describe the properties of such an environment and illustrate the perspective with existing related tools and environments.

Design novices have limited design experience and typically lack the skills or confidence to create good design, however, they may be able to recognize good design. To assess this ability, Kesler Tanner and James Landay had 53 novice designers and 52 expert designers participating in an online study where they evaluated a series of websites based on aesthetic appeal using two different modes of comparison. Results presented in ""I know it when I see it": How Experts and Novices Recognize Good Design" show that both experts and novices are able to recognize good design and that novices are able to do so almost as well as experts (76.5% accuracy compared to 81.2%). The greatest determinant of whether a participant would correctly identify a higher-rated design was the difference in the two websites' ground-truth aesthetic ratings. However, expertise and the mode by which the comparison was presented had a significant impact on accuracy (Keep-the-Best = 83.6% and Tournament = 74.1%).

3 Outlook

Extensive research conducted by the Hasso Plattner Design Thinking Research Program has yielded valuable insights on why and how design thinking works. Researchers have identified metrics, developed models and conducted studies that are featured in this book as well as in the previous volumes of this series.

We invite engagement with scholars of design thinking research for further discussion and an exchange of ideas. At www.hpi.de/dtrp you will find the latest information on all research conducted within our HPDTRP program, and learn more about its contributors.

Moreover, the website thisisdesignthinking.net offers an easily accessible overview of current developments in design thinking. This pool of examples and interviews, enriched with detailed explanations, helps to localize all existing expressions of design thinking, including their advantages and disadvantages. For educators, the website serves as a resource for clarifying explanatory models, and offering perspectives on current problems in design thinking practice. Experiences, stories and inquiries can be sent to thisisdesignthinking@hpi.de.

Through the dissemination of graduate-level research on design thinking, we aspire to produce a book series that become a preferred resource for informing future design thinking action.

Acknowledgements We thank all authors for sharing their research results in this publication. Our special thanks go to Dr. Sharon Nemeth for her constant support in reviewing the contributions. Thanks also to Chris Ford and Dagmar Willems for additional editorial support.

Theoretical Foundations of Design Thinking

Part II: Robert H. McKim's Need-Based Design Theory

Julia P. A. von Thienen, William J. Clancey, and Christoph Meinel

Abstract Although design thinking is often understood as a practical approach to creativity and innovation in design, it builds on highly refined theories. Many influential ideas were gathered and advanced at the Mechanical Engineering department of Stanford University from the 1950s onwards, as explored in this history series. In *part I* we introduced the "creative engineering" theory of Stanford educator John E. Arnold. This chapter—*part II*—is dedicated to the need-based design theory propounded by one of his successors at the department, Robert H. McKim. His theory, first published in 1959, advanced human-centred design conceptions by providing an elaborate account of human needs, by clarifying the role of designs and designers in the process of culture development, and by providing guidelines to assess, or actively increase, design value. According to McKim, the ultimate purpose of design is to promote the well-being of people by helping to gratify their basic needs. As his overall design framework is broadly scoped, it can also serve as frame of reference to analyse and compare different present-day approaches to design, such as innovation-focused design thinking and usability-focused studies in Human-Computer Interaction.

1 Introduction

As early as 1959, Stanford educator Robert H. McKim championed a design theory based on human needs that strived to go beyond the physical concerns of ergonomics to embrace a broader range of the human intellectual and emotional nature. McKim's theory is a precious resource for design thinking studies. First, he singles out and unites two subjects in John E. Arnold's varied teachings on creativity, which obtained a key role in design thinking many decades thereafter: design and human

J. P. A. von Thienen (✉) · C. Meinel
Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany
e-mail: Julia.vonThienen@hpi.de

W. J. Clancey
Florida Institute for Human and Machine Cognition, Pensacola, USA

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_2

needs. McKim's theory presents the two concepts so intricately connected as though they could not possibly be thought of independently.

Apart from the relevance of the topics, McKim's theory is also of historical relevance, as it set the course for notable subsequent developments. John Arnold had brought to Stanford a rich framework of creativity theories. It included the overall aim that designers satisfy human needs and the notion of a comprehensive designer, inspired especially by Buckminster Fuller's work. All along the way, John Arnold emphasized the importance of tangible results. In his framework, the creative process does not end with an abstract creative idea or insight, but with a tangible design. McKim elaborates on these thoughts and carries them further into practice. His articulation and teaching of product design and "rapid visualization" facilitated the move from design theory to design doing, and even today furthers a fruitful exchange between them. Thus, McKim establishes a knowledgeable and reflective culture of designers as "makers," emphasizing experience and prototyping, which has been a hallmark of design thinking up to the present.

While in many respects McKim's design theory elaborates ideas and ideals he shares with John Arnold, such as the key importance of human needs and self-knowledge, he also opens up novel perspectives on the subject that soon proved fruitful. Drawing on art theory and related fields of expertise, McKim proposed a conceptual framework with seeds of thought that he and his future colleagues subsequently developed into the rich notions of *Visual Thinking* and *Ambidextrous Thinking* (McKim 1972; Faste 1994).

Historically, it is also worth noting what elaborate design theories were already developed in the 1950s, which viewed design as a process of accommodating human needs. McKim presented a theory in scholarly style on this behalf. Arnold had integrated such ideas intricately in his creative thinking theories. And, of course, other protagonists like Fuller (1963) and Dreyfuss (1955) also advanced need-based design practices at the time. McKim's treatment of the subject comes particularly close to presenting a formal need-based design model—which, according to Hugh Dubberly's (2004) historically organized compendium, emerged in explicit form only a decade later. So, McKim's account can be particularly valuable to further elucidate the historical development of need-based design conceptions.

Yet another promising aspect of the theory reviewed in this chapter is the scope, which is very large, similar to Arnold's creativity theories. In Arnold's framework outstanding creative achievements can be treated as a special case, but the framework covers creativity in general, from minor to major achievements. Analogously, McKim presents a theory covering all instances of design, in which radical innovation can be treated as a special case. In this regard, his theory of design differs from present-day design thinking, which is often discussed specifically as an approach to design innovation (Brown and Katz 2009; Plattner et al. 2009; Meinel and Leifer 2011). The breadth of McKim's account supports using it as a more general framework, in which the relation of innovation-focused design thinking practices to other design approaches, such as usability studies in the field of Human-Computer Interaction, can be explored systematically.

This chapter figures as *part II* in the series *Theoretical Foundations of Design Thinking*. As in *part I*, in which we discussed John Arnold's creativity theories, we again endeavour to present the historical original in a systematic format, organizing the material in definitions (D), basic assumptions (A) and meta-theoretical remarks (M), to emphasize theoretical ideas that are particularly relevant for design thinking studies. The approach pursued here is clearly interpretative, notwithstanding many quotes from McKim's original text, especially because our summary is formulated from the perspective of design thinking research. We highly recommend reading McKim's (2016/1959) original essay, "Designing for the Whole Man," in addition to our analysis. It is available online in Clancey's (2016) *Creative Engineering* edition.

We will begin with a brief review of historical developments, tracing how McKim came to write the essay here discussed (section #2). The chapter then introduces McKim's definition of design (#3), his theory of human needs (#4), theory of culture development (#5), and theory of good design (#6). In the final section, we relate McKim's design framework to present-day design approaches, where we consider design thinking on the one hand and usability assessments, particularly in the field of Human-Computer Interaction, on the other (#7).

This chapter is specifically devoted to McKim's need-based design theory. His complete works go far beyond this subject. We expect to discuss further contributions he made in *part III* of the series *Theoretical Foundations of Design Thinking*, which will be dedicated to the concept of *Visual Thinking*.

2 The Emergence of Robert H. McKim's Essay on Design Theory

Robert H. McKim, born September 24 in 1926, moved to the Pratt Institute in Brooklyn to pursue Industrial Design after having earned a Mechanical Engineering degree at Stanford University. He came to work for Henry Dreyfuss in Manhattan as an industrial designer. Here, "human factors" were a prominently invoked concept. At the same time, McKim felt inspired to advance the human-centric perspective in design beyond what he witnessed in New York.[1]

The Dreyfuss studio in Manhattan in the late 1950s was a large room with a desk for every designer, fewer than a dozen altogether. Dreyfuss would personally tell every designer what human needs to consider; need-finding was thus a privilege of the design lead. Furthermore, core design ideas were often in place before human factors were considered, so these could only inform the style and packaging of a product, whereas the core design idea was developed independently. Finally, there was only one prototyper in the New York studio, working in the back of the room; none of the other designers would build prototypes.

[1] The following recollection was kindly shared by Bob McKim in personal conversation with William J. Clancey (12 Jan 2016, 16 Sept 2016, 19 October 2016, 31 Jan 2018).

McKim pondered a number of ways for advancing the human-centric perspective. He determined that, following Arnold's guideline "to understand man," core design ideas should emerge from a deep understanding of human needs. Furthermore, endorsing the notion of the "comprehensive designer," he envisioned the design process as less partitioned. Designers should be able to enact the whole process, including need-finding and building prototypes. Design for McKim is design doing, which in turn is a comprehensive process.

When McKim moved back to California, forming his own design firm in Palo Alto, he became aware of Arnold's work and planned to attend his *Creative Engineering* seminar in 1958/59 as a student.[2] They met ahead of the seminar, discovering an array of shared interests and visions. Both of them wanted to do "industrial design right," by which they meant focusing on human needs, designing "for the whole person" and educating students to become comprehensive designers instead of limited domain experts. In their first meeting Arnold suggested that McKim should attend the *Creative Engineering* seminar as a guest-lecturer, not as a student, which entailed writing an essay for the course manuscript. McKim agreed, and in his essay he formulated the need-based design theory reviewed in this chapter. Given the context of its emergence, unsurprisingly McKim's essay elaborates trains of thought that were of common concern to him and Arnold. Both of their accounts are included in the *Creative Engineering* manuscript dated 1958/59, which we will abbreviate as CE henceforth.

Besides McKim, Arnold invited other guest lectures to the course, who also contributed essays for the CE manuscript. The other guest authors were well-known then and now: the psychologists Joy Paul Guilford and Abraham Maslow as well as the philosopher Robert Hartman. This was the context and stage where McKim, at age 32 in 1958, developed and presented his design theory.

3 The Definition of Design

McKim straightforwardly presents a definition of the core concept of his need-based design theory.

(D1) "Design is the unique capacity of the human species to manipulate materials and energy in a reasoned or a felt response to human physical, intellectual, and emotional needs—human needs which are partially formed and modified by the natural and cultural environment" (McKim, CE, p. 200).

This definition indicates what phenomena McKim's theory sets out to clarify. In particular, he speaks not only about designed products, but takes specific interest in the process of their generation.

[2] Robert McKim, personal communication with William J. Clancey, 16 September 2016.

Fig. 1 Reconstruction of an early Neolithic Japanese pit house. (Image reprinted with permission from Aileen Kawagoe, originally published in *Heritage of Japan* (2018))

According to McKim, reasoned and felt responses to needs are approaches by which humans produce designs. This account parallels a key distinction in Arnold's framework, who speaks of two basic approaches for promoting creative problem solutions: *organized* and *inspired*. Relating Arnold's and McKim's conceptions helps to elucidate central ideas advanced from different angles by them and their fruitful interplay.

As Arnold explains: "The group of *organized approaches* is so named because they usually exhibit a logical, orderly, step-by-step type of problem solving technique" (Arnold, CE, p. 73, our emphasis). This group includes empirical trial-and-error procedures, as well as processes of careful reasoning and rationally deducing solutions. *Organized creativity approaches* re-appear in McKim's definition as *reasoned design responses*.

(D2) A reasoned design response is a non-instinctual way of addressing human needs, guided by a step-by-step rationale.

Interested in the origins of human design capacities, McKim gives an example from early hominin designs, indicating how these past creators where already capable of reasoned, that is, rationally planned design processes.

> This early Neolithic Japanese pit house (Fig. 1) looks very much like a nest which has been built on the ground by an enormous bird. A closer examination of the house and its contents would reveal, however, that its occupant has a very unusual talent for making a great variety of non-instinctual design responses. A reasoned design response, for example, makes possible the combination of a sharpened rock fastened to the end of a stick to make a hatchet with which to build the house. (McKim, CE, p. 198)

Here, McKim highlights the importance of rational planning to achieve the tools with which housing could then be built.

Inspired approaches form the second group of creative processes distinguished in Arnold's framework. He gives two examples. In what he calls the *Big Dream* approach (Arnold, CE, p. 67) people ask "big questions" or "dream the biggest dream they possibly can" and then spend ample efforts on answering the question or making the big dream come true. As another example of inspired procedures, Arnold describes the *Flash-of-Genius* approach. Here, the originator works tirelessly on a problem until an insight just pops into his or her mind in the form of a "eureka" experience, seemingly out of nowhere.

As the CE manuscript does not provide a straightforward definition, we tentatively defined inspired creativity approaches in our reconstruction of Arnold's framework as building "on intuition, fantasy or other loosely controlled psychological processes; they are characterized by relaxed ties to that which is considered possible, advisable or state of the art in the domain of creative work" (von Thienen et al. 2017, p. 29). What Arnold, in any case, makes very clear is the role and importance he attributes to inspired approaches.

> Inspired [. . .] approaches [. . .] are those closely associated with the art of creativity rather than the science. Big leaps in knowledge are apt to occur using these approaches, as compared with the slow but steady step-by step advancement made using organized techniques. (Arnold, CE, p. 73)

In McKim's theory, inspired creativity approaches re-surface as "felt design responses." He turns to art theory to further elucidate the process of creating designs by means of feeling, that is, how to enact an inspired creativity approach. In subsequent years McKim drew extensively on art theory and exercises from art education in the curricula he developed at Stanford.[3] In his CE essay he points out how discussions of artistic processes often make reference to emotions. By contrast, when people follow the orderly step-by-step approach of *reasoned design,* they are not necessarily relating to their feelings and experiences. Moreover, no reference to feelings seems necessary to evaluate a reasoned design. McKim invokes this observation to establish a defining contrast between reasoned/organized versus felt/intuitive design approaches. He suggests that feelings, or genuine experiences, play a key role in "felt design" processes, which he even names after this characteristic. The basic idea underlying this description McKim imports from art theory.

> As Clive Bell writes in his book, Art [1914, p. 50]:
>
> ["]To make the spectator feel, it seems that the creator must feel, too. What is this that imitated forms lack and created forms possess? [. . .] What is it that distinguishes the creator from the copyist? What can it be but emotion? Is it not because the artist's forms express a particular kind of emotion that they are significant?—because they fit and envelop it, that they are coherent?—because they communicate it, that they exalt us to ecstasy?["] (McKim, CE, p. 216)

[3]McKim epically advances a *Visual Thinking* curriculum, beginning with a course on Rapid Visualization in 1961, which is later re-named into *Rapid Prototyping*. An overview of courses offered by McKim is included in CE (Clancey 2016, p. 219f.).

Although in the CE manuscript from 1958/59 McKim seems to highlight the role of emotion, in line with Bell's original formulation, he will later on expand this perspective to experiences in general. The title of McKim's 1972 book states its objective: to help students have *Experiences in Visual Thinking*. This later curriculum covers much more than working with emotions. In particular, students are trained to use sensory experiences and imagery productively for creative purposes. To render the definition of felt design responses open towards these later developments, we suggest the following refinement:

(D3) A felt design response is a non-instinctual way of addressing human needs, guided by experiences.

In creative processes, experiences can play a guiding role in at least two ways. The first is in the form of emotional impulses such that some course of action may feel right, appropriate, exciting, promising, etc., while another can feel improper, unsatisfying, etc. To say that the person allows herself to be guided by what feels right, or to follow her urges and impulses, seems tantamount to John Arnold's description of people following intuitions in "inspired" creativity.

Yet, there is another sense in which experience can obtain a guiding role, and that is in the active seeking out of situations where the person has a chance to make and explore new ways of seeing, manipulating, feeling and thinking, ideally in ways that crystallize intuitions. This second sense of letting experiences guide the process will remain an important topic in McKim's work. In *Visual Thinking* (1972) he emphasizes how it helps to engage the senses, or the full body, in creative work. He explains how people could generally use different "thinking vehicles" (p. 3) to develop ideas. At school and the universities people were often trained to use verbal languages or mathematics. However, other thinking vehicles would also be highly valuable, and in this regard he especially points to "sensory imagery, and feelings" (p. 3), whose usage he trains in the classes he offers. *Part III* of this history series, dedicated to McKim's concept of *Visual Thinking,* will discuss these ideas and practices in further detail.

In his CE essay McKim already notes how reasoned and felt design responses need not come in pure form. Like Arnold, he also discusses combined cases.

(A1) Design responses can be both rational and felt.

Again, McKim provides an example from humanity's prehistory. He points to cave paintings, which obviously constitute artistic achievements (harnessing means of feeling), while requiring technical skill and preparation (drawing on means of reason) as well (Fig. 2).

> These cave paintings are a response, partly a reasoned response and partly a felt response, to this artist's intellectual and emotional needs to understand the mysteries of nature and to record his feelings about the world in which he finds himself. These drawings reveal that man is a good deal more than a reasoning creature with a unique ability for satisfying his own physical needs. He is, as well, a feeling creature with the ability to respond, by design, to emotional needs of a very high order. (McKim, CE, p. 199)

In this passage, McKim already refers to different kinds of needs, which shall now be discussed in more detail.

Fig. 2 Cave paintings as partly reasoned and partly felt design responses. (Image reprinted from KnowledgeNuts (2018); no image-copyright indicated)

4 A Theory of Human Needs

Although McKim sometimes invokes additional terms such as “visual needs” (CE, p. 214) or “instinctual [appetites],” (p. 213) his design theory specifically differentiates three groups of needs: physical, emotional and intellectual.

(A2) Physical needs include bodily well-being and the bodily aptitude to achieve desired ends.

As physical needs McKim explicitly mentions staying “alive, fed, and sheltered” (CE, p. 198) as well as “physical comfort and sensory well-being” (p. 217). In the latter case, typical design objectives are to avoid sensory unease or pain and to create circumstances where people can use and trust their senses. For instance, it is important that designs provide “illumination levels high enough and sound levels low enough” (p. 203).

Furthermore, people need to be physically capable of performing the actions and achieving the ends to which they aspire. This objective McKim discusses with an array of examples in which people’s physical needs are not met. For instance, kitchen machinery may be noisy and a sofa uncomfortable. Another example would be a garage door that closes automatically at such a rapid pace that some people fail to pass through quickly enough. Such designs do not accommodate people’s physical needs.

(A3) Emotional needs include experiencing positive or appropriate emotions and living out personal motives.

An arguably basic need is to experience positive emotions. McKim emphasises how designers can satisfy this need in straightforward ways. Designers should bear in mind “the delight which sensory stimuli such as color, shapes, rhythmic patterns, and textures can bring to the emotions” (p. 212).

Apart from the need to experience positive emotions, humans also need to have emotions that are appropriate to the situation. A fire alarm that strikes people as so charming that they want to linger and enjoy the sound would obviously be misconceived. Most of all, designs should “evoke emotions which are appropriate to the product” (p. 215).

Beyond basic or situation-specific needs, people also pursue personal motives, such as reinforcing preferred self-images. McKim quotes Vance Packard’s motivational consumer research: “A major appeal of buying a shiny new and more powerful car every couple of years is that it gives him (the buyer) a renewed sense of power and reassures him of his own masculinity, an emotional need which his old car fails to deliver” (p. 212).

Importantly, emotional needs or motives are different from emotions. Emotions convey needs of all categories.

(A4) All needs—including physical and intellectual—can be present to the person in the form of an emotional inclination, where she feels attracted or repelled by something.

Although some needs may be understood predominantly in terms of emotional motives, such as reinforcing a preferred self-image, feelings have an altogether mediating role in McKim’s framework. All needs can be—and usually are—present to the person emotionally. The narrow airplane seating that offends physical needs of moving the legs freely is experienced as unpleasant and the person wants to leave the inconvenient situation. Here feelings convey physical needs. Similarly, intellectual needs are experienced emotionally. “Purely intellectual appetites do not exist. To desire or to have an appetite for the ‘joys of the intellect’ implies an urge which is motivated by emotion” (p. 210).

(A5) Intellectual needs include understanding things and according with abstract ideals.

Understanding the purpose of tools or understanding messages quickly, operating tools with minimal intellectual effort, having designs accord to abstract ideals, and understanding nature are all examples of intellectual needs.

Generally, McKim notes, humans exhibit strong intellectual desires for knowledge; they even design theories about the world. It seems that already in humanity’s prehistory people experienced “intellectual [. . .] needs to understand the mysteries of nature” (p. 199).

Furthermore, intellectual needs inform the arts. Mondrian’s *Composition with Red, Black, Yellow, Blue, and Gray*, for example, “satisfies the intellectual appetites for unity, balance, and proportion” (p. 210).

In the domain of product design, McKim suggests two strategies to address intellectual needs: “(1) Minimizing needless intellectual effort required in the use of a product. (2) Satisfying the intellectual appetite for knowledge and order” (p. 204). Examples for achievements of the first kind can be found in jet aircraft designs, where it would be very dangerous if pilots puzzled over the meaning and function of cockpit knobs. “Human engineers have made extensive contributions to [. . .] design clarification, especially in the cockpit controls of jet aircraft where the minimizing of intellectual effort is essential to pilot safety.” Indeed, these designs could have safety implications much beyond the airplane crew. One does not want to imagine an army pilot who mistakenly releases an atom bomb while simply intending to switch on a reading light, all in confusion because his cockpit buttons were designed as look-alikes. The second objective is to satisfy intellectual appetites for knowledge and order. Examples for achievements in this regard obtain “when a design *looks* like it does what it does. [For instance:] A pleasing visual quality of some of our current automobiles is the fact they look like they are capable of going in a forward direction at a rapid rate of speed [which is exactly what they do]” (p. 205).

Notably, the three need categories provide heuristics for design; however needs by themselves can be ambiguous and sometimes include aspects of more than one kind.

(A6) The categories of physical, emotional and intellectual needs are not fully disjunctive.

Rhythm is one example for multifaceted needs, which McKim discusses drawing on Dewey’s elaborate treatment of the subject (cf. Dewey 1934). In McKim’s words: “Rhythm [. . .] is an aspect of design which is capable of satisfying a larger appetite, partly instinctual, partly physical, intellectual, and emotional.” Here, the “larger appetite” does not sort neatly into any singular category. Speaking of a “need for rhythm” or a “need for structure and repetition” could be rough approximations to capture this multi-faceted need.

5 A Theory of Culture Development

In his discussion of human needs and design McKim assumes an extremely broad historical perspective. Besides analyzing present-day conditions, he considers the evolution of needs and design in the course of human history.

Needs are described by McKim as highly dynamic. They continuously evolve in a co-evolution of needs and culture, where the latter includes designs. By tendency, the development of needs precedes and entails further advancements in human design, though the relationship is generally reciprocal. Developments on one side enable and entail further developments on the other, and vice versa.

(A7) Human culture, needs, and designs co-evolve.

As McKim lays out, in early forms of human life, ...

> needs are most often caused by some condition in the natural environment. The sun, rain, wind, the sea, the forest, animals in the forest—all had an enormous formative effect upon man’s early needs for design.

> But as man began to develop into the communal sort of life, into tribes and kingdoms, he soon found that he had to respond, by design, to an unnatural environment, which I shall call the cultural environment. (McKim, CE, p. 200)

(A8) Cultural developments bring about novel needs, which in turn bring about novel cultural developments, including designs.

> To illustrate the effect of the cultural environment upon human needs, and upon design, we may take as an example a recent design experience among many of the [. . .] tribes of Africa. For centuries, the warm natural environment of these tribes made the design response of clothing seem highly unnecessary. Their cultural environment was also quite untouched by the civilized values of the United States which require that men wear tight shirt collars, ties, and suits on a sweltering hot business day. But when Christian missionaries came upon the scene in Africa, a change in the cultural environment of many [. . .] tribes took place. The need for a design response—clothes—was experienced in very short order. Today, the native women are wearing calico dresses and the men are wearing dungarees, despite the hot weather. (McKim, CE, p. 200)

All in all, culture plays an ambiguous role in McKim's conception of human need and design evolution. On the one hand, culture is a driver of progress. Without cultural developments, human needs and designs might stagnate, or be simply reactive to happenstances in the natural environment. On the other hand, culturally induced concerns can be given priority over basic needs, engendering bad designs, unhealthy and unhappy people.

> In a modern society such as our own, the cultural environment probably has a more decisive effect upon human needs than does the natural environment. It often causes seemingly irrational needs for design which appear absurd to the people of other cultures. It causes fashions and styles in design. It sometimes frustrates the satisfaction of important human needs. But the design and art forms which constitute a good part of the cultural environment are the essential backbone of civilized values. It is a very stiff backbone, to be sure, but designers have, in the past, had remarkable success in bending it to their will. A cultural environment which frustrates the healthy satisfaction of human needs is, in my opinion, a culture which is in for a change. (McKim, CE, p. 200)

Here McKim describes culture-specific or culture-dependent needs that emerge, dynamically interrelating with physical, emotional and intellectual needs of mankind.

In more recent analysis, the developmental pattern McKim lays out can be reconstructed as a dependent hierarchy (cf. Wilden 1987; Clancey 2016). A level is called dependent on another, if it would disappear were the more fundamental level abolished. In McKim's framework, the physical level is the most fundamental one. Without physiological processes there would be no human emotions. As emotional creatures, humans also develop intellectual motivations and desires. Based on their intellectual capacities (including abilities of making reasoned design responses) mankind engenders a richly complex culture (Fig. 3).

All hierarchy levels are intricately interrelated, and causation can work bi-directionally. When a person becomes physically ill, her emotions, intellectual capacities and opportunities to produce or enjoy culture may be affected. Conversely,

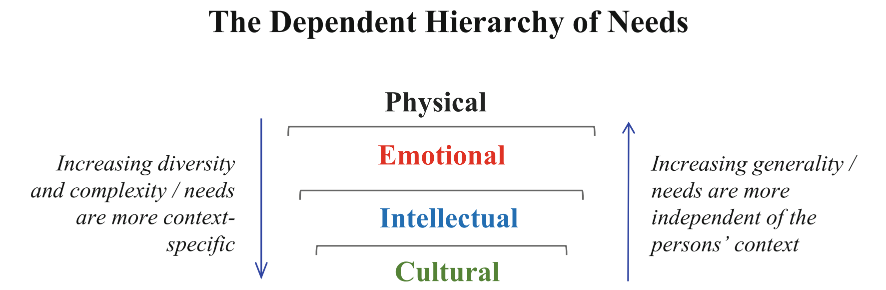


Fig. 3 McKim's discussion of needs that unfold on the levels of physical, emotional, intellectual and cultural requirements, or motives, can be analyzed in the form of a dependent hierarchy

an unjust cultural environment can affect the intellectual processes of community members; it may elicit emotional upset and thus even physiological impairments.

McKim describes the evolution of needs and culture in the course of human history as a process, in which the impact of culture increased over time. Thus, increased rates of activity and downward influence would be expected on the cultural level if the evolving dependent hierarchy were to be modelled chronologically.[4]

Notably, in his essay McKim elaborates on three groups of basic needs—the physical, emotional and intellectual—while not refining what "basic" means. Based on his theory we suggest the following tentative definition.

(D4) Basic human needs have already existed in humanity's pre-history; they hold cross-culturally and must be gratified regularly in order to ensure long-term health and happiness.

While the basic needs are taken to be particularly important for human designs and the objective of human well-being, people are taken to actively determine the gratification priority they assign to needs. It can happen that a culturally-induced, non-basic need obtains a higher gratification priority than a basic need. This is not only a matter of individual choice; society or one's reference group can prioritize some needs over others.

(A9) Both individuals and larger groups can develop distinctive preference patterns were gratification priority is assigned to some needs rather than others.

[4]Reconstructing McKim's ideas in the form of a dependent hierarchy explores their potential to be understood in terms of a single, unified and complete hierarchy of human needs. In his CE essay, McKim suggests no particular ordering among the physical, emotional and intellectual need categories. The term "need hierarchy" subsequently used also does not presuppose any such order. It refers to the possibility of needs being ordered from more basic/general to more concrete/context-specific/culture-dependent. Resulting need hierarchy branches can remain within a single need category. E.g., in the physical need spectrum a person may need... "to stay healthy...to endorse physically healthy postures... to have more head-room in her automobile".

Fig. 4 Culture evolves as novel needs promote the development of novel designs. These become part of culture, so that the cultural environment of people changes. Finding themselves in novel situations, people develop novel needs. They also change their need hierarchies, re-evaluating the gratification priority they assign to needs

On these grounds, McKim describes an endless, iterative process of culture development. Novel designs bring the community into novel situations. This may lead to an updating of the need hierarchies, with gratification priorities differently assigned. In addition, novel needs may emerge, which in turn prompt designers to develop new artefacts. So the cultural environment changes etc. Figure 4 visualises the process of culture development McKim suggests.

Against this background of culture development, McKim discusses the role of designers. He highlights how they are important contributors in the process of culture formation. Indeed, they have the power to change patterns of need-gratification.

(A10) Designers shape culture for good or bad. It is the designer's responsibility to understand and, if necessary, influence the cultural environment, to satisfy the basic needs—to design for the whole person.

We provide a tentative definition of the term "designer" to underpin the framework. McKim did not spell out an explicit definition himself, but he clearly does not reserve the term "designer" for people living in recent times. Pre-historical humans who built Neolithic pit houses and engendered cave paintings are addressed by McKim as designers. In his framework they even exemplify what it means to be a "human designer" (p. 200).

We suggest a twofold refinement to explicate McKim's outlook on designers: a more liberal and a stricter definition. The more liberal account is intended to capture pre-historical cases, in which little is known of people's reflective abilities. The stricter definition adds one criterion in brackets, requiring of designers increased self-awareness and anticipation.

(D5) A designer is an individual who—by means of reason and/or feeling—develops a new tangible or non-tangible product to address needs (and who anticipates a reproduction of the product to address needs of the same kind repeatedly in the future).

McKim discusses the activities of present-day designers in terms of the stricter definition. He adds that designers should not only anticipate the reproduction of their designs, but ought to envisage also the ensuing effects. The automobile industry provides critical examples.

> Jagged door openings, insufficient head, leg, and knee room, and uncomfortable sitting positions are but a few of the unpleasant physical features of the latest "advances" in automobile styling. Of course, automobile manufacturers know full well that they are offending the human anatomy with their design. They claim, and they have reams of market research statistics to back them up, that the public prefers the long, low, fast look to being comfortable. Comfort, in other words, is an insignificant need in comparison to the emotional needs which are satisfied by current modes in styling. (CE, p. 202)

In this example it is not only an individual, but "the public" who endorses a non-holistic design. Priority is not placed on the satisfaction of all the basic needs, including the adoption of physically healthy postures, but on gratifying emotional needs of driving fashionable, long, low cars. Designers in the 1950s supported unfavourable patterns of need-gratification in society by providing ever more emotionally-attractive, uncomfortable cars.

(A11) Culturally induced needs can obtain a high gratification-priority both in human desires and designs, leading to the harmful neglect of some basic human needs, causing unhappiness and illness over time.

Reconsidering these objectives in McKim's design treatise from the present-day perspective of design thinking research, the relationship between design(ers) and culture can be identified as a topic of enduring interest. Recently Jonathan Edelman pondered how "designers create culture" (2017a, p. 8). Barry Katz (1990) added another perspective by exploring also the other way around, how culture impacts the persons who develop novel artefacts. He observed how some cultures promote innovation while others prefer and generate more tradition-bound designs.

M1: Design thinkers continue to research the relationship between design, designers and culture.

Based on his model of cultural evolution—where developments can generally occur for the good or bad of humanity—and in light of the important role he attributes to designers, McKim spells out an explicit theory of good design.

6 A Theory of Good Design

Unlike natural occurrences that simply happen one way or other, humans develop their designs purposefully. McKim emphasises how "design is ultimately for the well-being and happiness of man" (CE, p. 198). Generalising this formulation just a little to avoid misinterpretations, such as an exclusion of women, the purpose of design might be defined as follows.

(D6) The ultimate purpose of design is the well-being and happiness of creatures, humans in particular.

In order to achieve that, McKim emphasises how designs must address human needs *comprehensively*. "As in a well-balanced life, good design must balance the requirements of physical, intellectual and emotional needs" (CE, p. 211).

Addressing needs comprehensively also means that in each need category a great bandwidth of different concerns must be accommodated. For instance, concentrating on only one need in the physical realm—that of amplifying human powers—is unlikely to spawn good designs.

> The human values in design which are the chief concern of modern technology are physical values. The engineer and scientist are primarily concerned with extending man's physical power over his environment. Utopia, to the engineer, would be a world in which the most strenuous physical task would be the pushing of a button. This Utopian vision, which presumably has the majority of the world's population sitting at home contemplating its navel, is perhaps not too far off. (CE, p. 200)

Such a utopian vision could be realised with machinery that is physically offensive to the senses and designs that induce unhealthy postures as well as too little physical activity. To achieve a good design, in each need category the full spectrum of basic needs must be accommodated.

Furthermore, it is important that each design covers needs from all three categories—physical, emotional and intellectual—instead of focussing selectively on needs of just one kind. In this regard, McKim especially criticises an "overemphasis, in modern design, upon intellectual needs" (p. 210).

> In my opinion, "good design" and the slogan "Form Follows Function" has come to mean an almost exclusive emphasis on the intellectual values of visual clarity of function, structure, and materials plus the visual application of the intellectual principles of unity, balance, and proportion.
>
> It is certainly all to the good for design to satisfy the thinking man. But what of the function of design, in terms of other human values? Man also has emotional needs. Man has shown an irrational appetite for decoration, for example, since his earliest utilitarian art—can modern design revolutionize this human need for decoration out of existence? Man has also a great irrational need for being enclosed, cozy, secure [...]—can modern design with its gold-fish-bowl expanses of glass, liquidate this human need for security? (McKim, CE, p. 211)

As a final requirement for good design, McKim invokes a moral evaluation of the needs that designers set out to gratify. The designer's goal cannot be to satisfy as many needs as possibly in each category, for some human desires are actually immoral.

> Not all emotional needs are good ones. The bloody history of the weapons of war is but one example of man fulfilling the wrong emotional needs through design. When we consider designing for the emotional needs of man, therefore, we come quickly to the question of morality in design. (McKim, CE, p. 211f.)

In this regard, McKim is especially concerned that "motivational researchers have spurred designers on to satisfy emotional needs that would perhaps be better left unsatisfied" (CE, p. 216). Building and selling ever faster cars that appeal to the buyers by means of conveying power and masculinity are cited as one example in place.

> With 33,000 dead and 5,000,000 injured as a result of last year's automobile accidents, this design response to emotional needs for "power" and "masculinity" seems to me to be decidedly immoral. Our morality of behaviour certainly does not sanction everyone to go around carrying a loaded gun so that they can be "reassured of their masculinity." It seems to me that designers should have similar standards for design morality, standards which would be applied at the inception of every new design. (McKim, CE, p. 212)

(D7) A good design accommodates (1) basic needs in all three categories: physical, emotional and intellectual, (2) comprehensively addressing relevant needs in each category, while (3) satisfying civilized needs only.

These are necessary but not sufficient conditions; a design must of course also satisfy context-specific requirements, such as situation- or culture-specific needs.

As one technique to elucidate the moral value of a design, McKim suggests to imagine the use of the designed product and analyse the ensuing behaviour.

> Today it is considered immoral (and illegal) to use a switchblade knife; but you may design one without restraint. It is moral to design a thermonuclear weapon; fortunately, it is currently considered immoral to use one. The emotional values of many of our most popular products, if translated to human behaviour patterns, would certainly seem, if not immoral, at least overbearingly arrogant and power-happy. (CE, p. 212)

(D8) A designed product satisfies civilised needs when typical forms of product usage entail morally acceptable behaviour.

From a meta-perspective, McKim's move to weave normative questions of design value and even ethics right into his theory of design is a noteworthy one. In design thinking research quite disjunctive work traditions have been identified that either focus on descriptive objectives or address normative questions as well (Lindberg et al. 2009). The meaning that is associated with the term "design thinking" differs considerably up to the present across these different work traditions (Lindberg 2013). Notably, openness to normative concerns has been highlighted as a special characteristic of the design thinking approach pioneered at Stanford.

> Advanced mainly by research from Brian Lawson and Nigel Cross, 'Design Thinking' was initially communicated as an open concept to describe cognitive problem solving strategies in design processes [studying descriptively what design experts actually do]. Subsequently, in the 1990s, a normative-pragmatic conception of design thinking became increasingly known, which was developed at the Stanford School of Engineering. It was used both in business practice (e.g., IDEO, frog design) as well as in the form of didactic concepts developed at the schools of Design Thinking (d.schools). (Lindberg et al. 2009, p. 47, our translation)

This concern for normative questions that Lindberg et al. highlight to be a special characteristic of Stanford's design thinking approach as advanced from the 1990s onwards, can now be seen to have much longer historical roots. To McKim, normative questions inform the heart of design. He naturally treats these questions in his design theory and encourages the readers to reflect on them.

(M2) McKim's need-based design theory is both descriptive and normative.
(M3) On normative grounds, McKim addresses what designers should do, such as gratifying basic needs, and what makes a good design.

With this discussion we end the introduction of McKim's need-based design theory and now explore its relations to present-day design approaches, especially in the realm of design thinking and Human-Computer Interaction (HCI).

7 Relations of McKim's Framework to Design Thinking and HCI Usability Research

As McKim's need-based design theory was formulated in the 1950s, a considerably long time ago, a likely reflective question to conclude this chapter addresses relations to present-day design approaches, in particular design thinking. To what extent did McKim's ideas anticipate or go beyond recent-day practices? What concerns have endured or changed? As will soon become evident, strong continuities can be observed from his theory to present-day design thinking. Yet, in some regards the need-based design theory also goes beyond recent design thinking concerns. Some of McKim's ideas rather seem continued in usability studies, as advanced in the field of Human Computer Interaction.

On behalf of design thinking, a first thing to note is the enduring concern for basic human needs. They shall inform core design concepts, instead of being merely invoked as styling rationales. Products developed by design thinkers are directly tailored towards the careful and intentional gratification of important human needs. That is why design thinkers learn to empathize with users prior to thinking up designs. When interviewing a user, design thinking students are told,

> we want to understand a person's thoughts, emotions, and motivations, so that we can determine how to innovate for him or her. By understanding the choices that person makes [. . .], we can identify their needs and design for those needs. (d.school 2010, p. 10)

As design thinkers are well aware, users often report context- or person-specific concerns when asked straightforwardly about their needs. *Why-How-Laddering* is invoked as a technique in present-day design thinking projects to make sense of user needs in a more systematic fashion (d.school 2010). Here, needs are analysed in the form of need hierarchies. They are sorted from more concrete, context- or person-specific to more basic, context-independent and common-human.

> When considering the needs of your user, start with a meaningful one. [. . .] Ask why your user would have that need, and phrase the answer as a need. For example [looking at the handling of food], "Why would she 'need to see a link between a product and the natural process that created it'? Because she 'needs to have confidence that something will not harm her health by understanding where it came from'." [. . .] At a certain point you will reach a very abstract need, common to just about everyone, such as the 'need to be healthy'. This is the top of that need hierarchy branch. (d.school 2010, p. 20)

This understanding of needs resembles McKim's in that needs can be elaborated in a dependent hierarchy from basic/common-human to concrete/context-specific. In McKim's framework, similar need hierarchy braches are hinted at, for instance, in the emotional need category. Men who cherish powerful sports cars may be described as needing "…to feel socially respected …to feel masculine and powerful …to show off with a powerful car".

Clearly, merited designs in terms of human values emerge when the more basic needs are addressed. Needs or desires that are relatively more concrete, context- or person-specific can be immoral or power-happy, such as wanting to drive venturously with a powerful car; a socially responsible designer would not want to support these kinds of needs.

On behalf of need-terminology, a minor refinement has been developed after McKim formulated his theory. Today students learn: "Needs are *verbs* (activities and desires with which your user could use help), not *nouns* (solutions)" (d.school 2010, p. 15). McKim does not yet invoke this refinement. He speaks of a need for clothing (even in particular styles), decoration, comfort, delight, power etc., thus phrasing needs in terms of nouns. Just as regularly, he uses verbs, describing people as needing to stay alive, fed and sheltered, needing to know or understand something, needing to be cosy or needing to record their feelings.

Yet, behind superficial differences in the articulation of needs lies an important continuity. The term "need" is used in a permissive way. Essentially all culture-, situation- or person-specific requirements and desires can be rendered as needs, just like the common-human concerns to stay alive, to have something to eat or to be socially accepted. Only due to this permissive term usage does it make sense to evaluate needs morally, as McKim does. The same holds for the designer's self-selected task to differentiate between needs that shall be addressed by design and those that remain unattended—a selection task clearly articulated in McKim's theory and accepted as a matter of course by design thinkers at present. If, by contrast, the term "need" was used in a stricter sense, reserved only for common-human objectives such as staying alive, no moral evaluation would be possible and designers should better attend to every human need.

M4: The term "need" is used permissively by McKim and present-day design thinkers. By implication it refers to all human concerns, from common-human objectives such as staying alive ("basic needs") to culture-, situation- or person-specific objectives.

M5: The designers' task to organise needs in hierarchies, sorting them from "more basic" to "more concrete", and the task to evaluate which needs shall be gratified by design, derive from permissive talk about "needs".

M6: Design thinking continues McKim's human-centred approach of establishing good designs by envisioning relatively more basic needs behind concrete, context- or person-specific objectives.

Next to these strong lines of continuity, some relevant changes can also be observed from McKim's theory to present-day design thinking. Notably, McKim does not refer to "innovation" even once in his essay, while present-day design thinking is typically understood as an approach to engender design innovation. Meinel and Leifer articulate this objective clearly in their description of the "first two schools of Design Thinking" at Stanford and Potsdam University.

> "We believe great *innovators* and leaders need to be great design thinkers. We believe design thinking is a catalyst for *innovation* and bringing new things into the world. We believe high impact teams work at the intersection of technology, business, and human values. [...]" These are the visions of the first two schools of Design Thinking [...]. With overwhelming success these schools educate young *innovators* from different disciplines [...] to work together to solve big problems in a human centered way. (Meinel and Leifer 2011, p. xiii, our emphasis)

M7: McKim's design theory has a broader scope than present-day design thinking in the sense that it covers all instances of design, while design thinking is specifically concerned with innovative design.

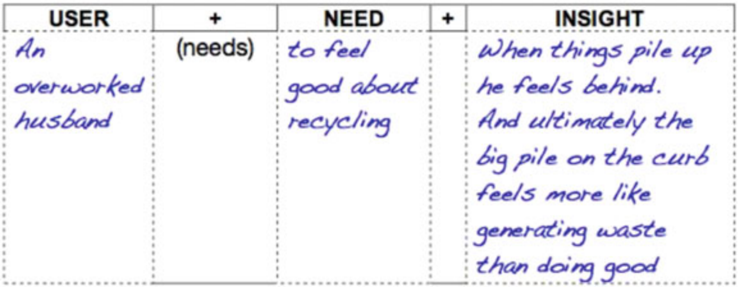
However, to avoid misconceptions, it should be recalled how Arnold's creativity theories were concerned with innovation (Arnold 1959/2016; von Thienen et al. 2017); and Arnold clearly interpreted McKim's need-based evaluation as contributing to the creative engineering curriculum. So this objective of present-day design thinking has long historical roots as well, arguably including McKim's need-based design theory.

Also, with regard to the design process McKim's theory can be said to have a broader scope than present-day design thinking. McKim's theory discusses design processes comprehensively. Human values are assigned a key role throughout, from the inception phase, where he suggests that human factors ought to inform the core design concept, up to final design decisions, where he prompts designers to be aware of the impact that colours, texture and knob positions have on the user. Crucially, the theory underscores a multiplicity of needs that each design must ultimately address on all levels of analysis, the physical, emotional and intellectual. These are objectives that come into play when designs undergo finalization.

By contrast, according to present-day design thinking such details do not yet matter in the inception phase. It is instead considered helpful to concentrate on one single basic need to develop the core design idea.

In design thinking methodology, the search for basic design ideas is directed by a *How-Might-We* question, which is structured as [user] + [need] + [insight]. In the "need field" one does not state an array of physical, emotional and intellectual needs, but concentrates on a single objective (Figure 5 provides an example).

Based on the *How-Might-We* question, design thinkers gather multiple potential solutions. One core design idea is then selected, prototyped and tested. Again, in the prototyping phase design thinkers often do not work with highly refined artefacts that would address the full range of physical, emotional and intellectual needs with carefully chosen knob positions, textures and colours or broader considerations of safety, comfort, and usability. Design thinking research has pointed out the benefits of working with rough prototypes (cf. Fig. 6), which promote open discussions about core design ideas and allow different concepts to arise. By contrast, highly refined



USER	+	NEED	+	INSIGHT
An overworked husband	(needs)	to feel good about recycling		When things pile up he feels behind. And ultimately the big pile on the curb feels more like generating waste than doing good

Fig. 5 Design thinkers learn to concentrate on one need to be addressed by design. It is specified in the *How-Might-We* question, which directs the search for solutions (Creative Commons image reprinted from the bootcamp bootleg, d.school 2010)

prototypes with many design details already in place seem to stimulate conversations about solution details, not discussions of basic solution approaches (Edelman and Currano 2011).

Building on such research insights, design thinking teams are encouraged and schooled in the production of rough prototypes. Teams can use them as a tool to learn whether or not they are on a good track with their focus on a particular user need, and with their general idea of how to gratify the need with a novel design. By contrast, clarifications of precise design details are usually thought to come after a design thinking project, for example, when the core design idea is handed over to the production department in a big company. Present-day design thinking models concentrate on initial phases in a design process, whereas later phases—getting all the details right, moving on to production, promoting an innovation despite of potential resistance and criticism—do not lie at the focus at present (von Thienen and Meinel 2014, 2015).

M8: McKim's design theory has a broader scope than present-day design thinking in the sense that it covers the overall design process, from the inception of a core design idea to the finalization of all design details, while present-day design thinking focuses on the inception of a highly valuable and innovative core design idea.

In accord with its broader scope, McKim's design theory also submitted ideas for the finalisation of designs that historically came to be advanced in other design approaches, apart from design thinking. In this regard it is especially noteworthy how his theory anticipated concerns and methods that would later be elaborated under the rubric of *usability* in the field of Human-Computer Interaction (HCI).

The HCI approach to design bears some resemblance to design thinking insofar as a human-centred perspective is endorsed and ample tests are launched to ensure good designs. There is furthermore a joint focus on technology. HCI explicitly refers to computers, while design thinkers are recognized for working at the intersection of "*technology*, business, and human values" (Meinel and Leifer 2011, p. xiii, our emphasis). None the less, relations of the two approaches are seldom clarified. Given

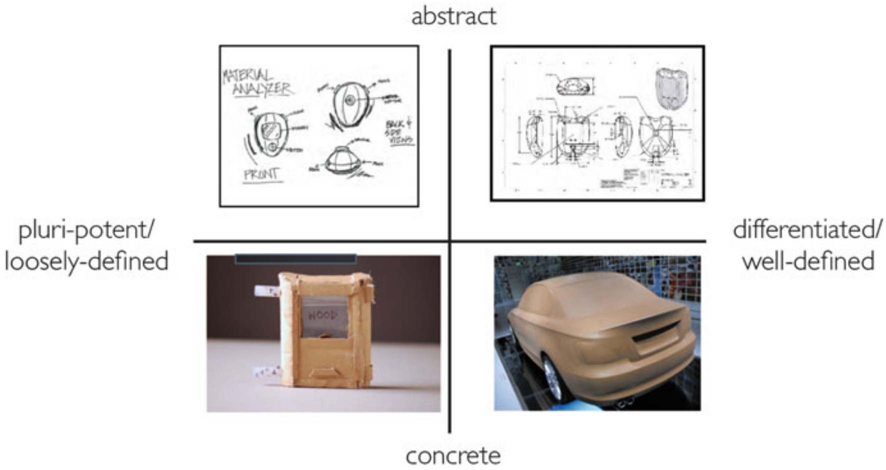


Fig. 6 The media model framework (Edelman and Currano 2011), developed in design thinking research, sorts prototypes according to their level of refinement and abstraction. Design thinkers often, intentionally, work with unrefined prototypes (left side of the matrix). These help to communicate about core design ideas. By contrast, refined prototypes (right side of the matrix) are suggestive of almost finalized solutions; they rather stimulate conversations about fine-grained design details (Image adapted with permission from Edelman 2017b; including the depiction of a clay car model by Bigge, CC BY 3.0)

that design thinking is innovation oriented, while HCI is neutral in this regard, one could imagine merging the approaches by incorporating a design thinking perspective during HCI's inception phase and using unrefined prototypes.

This chapter shall close with an exploration of ideas in McKim's theory that seem to continue in usability (more than in design thinking) studies at present. Here, McKim's theory can serve as a systemising framework, in which the relation of design thinking and HCI can be further clarified. Moreover, the discussion brings us back to McKim's original formulations, with seeds of thought that design thinkers might wish to pick up (again) in the future. The potential of the need-based design theory to underpin design thinking practices might not be fully realised as of yet; further ideas could be re-integrated and elaborated.

In the field of *Human-Computer Interaction* (HCI), Dix et al. (2004) provide a dense introduction to the subject of usability. They name three objectives that designers ought to bear in mind to achieve high levels of usability. Designers should attend to the objectives of...

> **Learnability**—the ease with which new users can begin effective interaction and achieve maximal performance.
>
> **Flexibility**—the multiplicity of ways in which the user and system exchange information.
>
> **Robustness**—the level of support provided to the user in determining successful achievement and assessment of goals. (Dix et al. 2004, p. 260)

McKim's design treatise was clearly concerned with related objectives. He can be said to have anticipated *learnability* maxims as he thought about the time and effort people had to invest when trying to handle some product for the first time. "We have all had the frustrating experience of not being able to understand how to turn a simple product on or off. Clairvoyance is certainly required to divine that the rotating knob on several popular appliances must be *pulled out* to turn the appliance on" (McKim, CE, p. 204, emphasis in original). He also offers advice that could just as well appear in a present-day usability textbook:

> Many puzzlers could be solved with greater ease if less reliance were placed on instruction manuals and more thought given to "building in" the instruction by means of design clarification. Modern kitchen stoves, for example, are often partially inoperable without complex instructions from a manual. Many of these instructions could be built into the design in the form of the logical arrangement of the controls into flow patterns which visually indicate operating procedure. (McKim, CE, p. 205)

In a similar vein, he advocates for hierarchies in operating controls to make the product easily usable—what in present-day terminology we would address as an aspect of *learnability*. Dominant design elements such as big, colourful buttons, should be reserved for important functionality.

> A "functional hierarchy" for a low-priced TV set might be established, in terms of a broad "human use factor," as: 1. Picture tube [i.e. screen in present-day TVs]. 2. Channel selector. [. . .] 6. Speaker. [. . .] To arbitrarily reverse this hierarchy in the visual ordering of these components would normally be quite undesirable. For example, to satisfy an "aesthetic impulse," the designer might make the volume control overly prominent in terms of its color, size or position—thereby distracting the eye from No. 1 of the functional hierarchy—the picture tube. (McKim, CE, p. 208)

Issues of *flexibility* are also addressed straightforwardly in McKim's framework. He especially criticises designs of his time that forced users to rigidly use one specific interaction mode, relying on a single sense-channel.

> Every evening thousands of Americans climb into their automobiles, reach for the headlight knob, turn instead its identical twin, the windshield wiper knob, or perhaps its triplet, the cigarette lighter. It is not difficult to find examples of "Chinese puzzles" in our everyday design world. Unfortunately these puzzles are not fun; they are frustrating. (McKim, CE, p. 204)

He underscores how there should be multiple ways of interacting with a device. In addition to visual clues, there could be tactile ones.

> The headlight—windshield wiper— puzzler [. . .] could easily be minimized in several ways.
>
> (1) *Coding* the knobs by shape or texture so that their differences would be tactually clear—day or night. The confusion that arises with these controls usually takes place when it is dark.
>
> (2) *Positioning* the knobs according to their respective functions—the windshield wiper knob near the wipers, the headlight knob near the ignition key for handy use when starting up at night. (p. 204f.)

All in all, designers are prompted to consider differing usage scenarios, such as driving at day versus at night. Designers should offer multiple ways for users to operate a device, people should be flexible in how they interact with a machine.

The issue of *robustness*, in turn, is not to be found in a single quote in McKim's essay, though it is arguably a central concern throughout. Nevertheless, McKim's focus is somewhat different from present-day usability research about computer systems that engage people in prolonged, sequential or recurrent interactions.

In HCI a quick recognition of, and recovery from, errors has become of notable concern (e.g., Nielsen 1994). Questions strictly similar to those raised in many HCI studies—such as what happens after a mistaken mouse click, how long it takes to find the undo button or another way around—are not prominently discussed in McKim's design theory. However, it is clear that not being able to recover quickly from a mistake is frustrating, and user frustration to McKim is an unambiguous indicator that the device needs to be redesigned.

The aspect of robustness that is indeed centrally discussed in McKim's theory (as well as in Arnold's framework) amounts to finding out what goals the users *really* pursue, what needs people *really* have. Designers who develop technology are warned of the perils of automation and modernisation for their own sake. In one typical example McKim reviews a "modern" house and its kitchen machinery.

> The house itself, a collection of all the geometrical clichés of modernism, was the essence of cold impersonality. The automatic kitchen was a nightmare of whining motors and flashing lights. (McKim, CE, p. 198)

On behalf of present-day usability research Dix et al. (2004) explain: "In a work or task domain, a user is engaged with a [...] [product] in order to achieve some set of goals. The robustness of that interaction covers features that support the successful achievement and assessment of the goals" (p. 270). McKim would highlight how designers need to understand "the user's true goals", that is, their basic needs, in order to design suitable products for them. From McKim's point of view, greatest robustness hurdles might not stem from clumsy recovery processes, but from mismatches between the goals a product is designed to support and the goals users actually pursue.

M9: McKim's theory anticipates design maxims that are advanced under the headline of usability in the field of Human-Computer Interaction; they bear on the *learnability*, *flexibility* and *robustness* of user-product interactions.

What HCI usability research and McKim's design theory also share is their sensitivity to a huge array of final design details. To assess a product's usability from an HCI perspective, a large number of functional design issues must be settled first. Only then can typical human factors questions be answered: whether all operating procedures are internally consistent, whether the use of colour and sound is not distracting, whether the chosen language labels are understandable etc. Similarly, McKim requires of a good design that it satisfies multiple physical, emotional and intellectual needs. How well a design figures in this regard depends not only on the core design idea. Rather, McKim emphasises, it is highly important what forms, materials, colours, textures, etc. the designer selects.

M10: McKim's design theory and usability studies include assessment strategies of design value that work for refined artefacts only; they require that final design details are already in place.

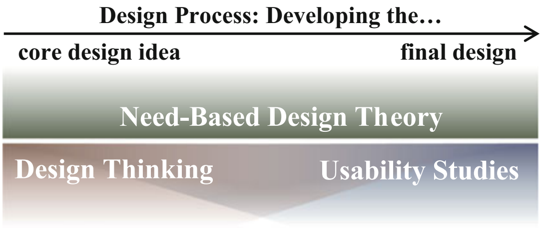


Fig. 7 Design thinking focuses mostly on the development of a valuable and innovative core design idea, while fewer resources are devoted to the clarification of final design details. Usability studies are specialized on the clarification of final design details, while the process of developing a core design idea is less elaborated. McKim's need-based design theory spans the overall process

M11: Given its broad scope, including also accounts of design value, McKim's theory can serve as an overarching framework, in which different present-day design approaches with varying objectives (focusing on design inception versus finalization, on innovation vs. general usability etc.) can be systematically compared.

All in all, a broad-scope design theory is offered by McKim, which covers the design process from early to late stages, consistently from the point of view of human needs (Fig. 7). Even in late design stages—where details are attended, such as colours, texture, sound-levels, the size and positioning of knobs etc.—McKim's design theory consistently analyses designs and human-product interactions from the perspective of user needs.

The passage that closes McKim's CE essay shall also close this review of his theory. It reflects a great continuity of concern for truly valuable designs, to be achieved by working in a human-centred way, by addressing human needs comprehensively. McKim states the objective in obvious synergy with Arnold's visions and clearly anticipating present-day design thinking practices. While many of the submitted ideas have already been picked up and are fruitfully elaborated in present-day design thinking culture, McKim's overall framework offers a cornucopia of suggestions that can continue to provide inspiration.

> Clearly we badly need the designer who understands and is capable of responding to the needs of the whole man. This designer should be capable of reasoned as well as felt design responses. He must understand man's physical needs, needs not only for power over his environment but needs for physical comfort and sensory well-being. He must understand man's intellectual needs, needs for minimizing needless problem solving in design as well as visual needs for knowledge and order. The designer who designs for the whole man will also understand man's emotional needs for designs which satisfy civilized motivations and which delight the emotions through the senses. This designer must have the fortitude to exert his influence on the current cultural environment which is depriving us all of basic human needs. (McKim, CE, p. 217)

Acknowledgements We thank Robert H. McKim for his interest and support in the reconsideration of his need-based design theory. We thank Anja Perlich and Jonathan Edelman for helpful discussions and are grateful to Aileen Kawagoe as well as Jonathan Edelman for the permission to re-print their images.

References

Arnold, J. E. (2016). Creative engineering. In W. J. Clancey (Ed.), *Creative engineering: Promoting innovation by thinking differently* (pp. 59–150). Stanford Digital Repository. Available from http://purl.stanford.edu/jb100vs5745 (Original manuscript 1959).

Bell, C. (1914). *Art*. London: Chatto and Windus.

Brown, T., & Katz, B. (2009). *Change by design: How design thinking transforms organizations and inspires innovation*. New York: Harper Collins.

Clancey, W. J. (Ed.). (2016). *Creative engineering: Promoting innovation by thinking differently, by John E. Arnold. Edited with an introduction and biographical essay by William J. Clancey*. Accessed February 2018, http://purl.stanford.edu/jb100vs5745

d.school. (2010). *Bootcamp bootleg*. Accessed February 2018, from http://longevity3.stanford.edu/designchallenge2015/files/2013/09/Bootleg.pdf

Dewey, J. (1934). *Art as experience*. New York: Perigee.

Dix, A., Finlay, J., Abowd, G. D., & Beale, R. (2004). *Human–computer interaction*. Harlow: Pearson, Prentice Hall.

Dreyfuss, H. (1955). *Designing for people*. New York: Simon and Schuster.

Dubberly, H. (2004). *How do you design? A compendium of models*. San Francisco: Dubberly Design Office.

Edelman, J. A. (2017a). *Dimensions of engagement*. HPI Brown Bag Series, December, 2017.

Edelman, J. A. (2017b). *Bridging the research//practice gap: Towards innovation algorithms and qualitative functions in domain specific design scenarios*. HPI D-Flect, December 2017.

Edelman, J. A., & Currano, R. (2011). Re-representation: Affordances of shared models in team-based design. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking. Understand – improve – apply* (pp. 61–79). Heidelberg: Springer.

Faste, R. (1994). Ambidextrous thinking. In *Innovations in mechanical engineering curricula for the 1990s*. New York: American Society of Mechanical Engineers. Accessed February 2018, from http://www.fastefoundation.org/publications/ambidextrous_thinking.pdf

Fuller, B. (1963). *Ideas and integrities: A spontaneous autobiographical disclosure*. Englewood Cliffs: Prentice-Hall.

Katz, B. (1990). *Technology and culture: A historical romance*. Stanford: Stanford Alumni Association.

Kawagoe, A. (2018). *Heritage of Japan* [website]. Accessed February 2018, from https://heritageofjapan.wordpress.com/just-what-was-so-amazing-about-jomon-japan/ways-of-the-jomon-world-2/jomon-architecture/

KnowledgeNuts. (2018). *The strange truth about the people who painted cave art* [website]. Accessed February 2018, from http://knowledgenuts.com/2014/04/04/the-strange-truth-about-the-people-who-painted-cave-art. Article published April 4, 2014.

Lindberg, T. S. (2013). *Design-thinking-Diskurse: Bestimmung, Themen, Entwicklungen.* Doctoral dissertation, University of Potsdam, Germany. Accessed February 2018, from https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/index/index/docId/6733

Lindberg, T. S., Noweski, C., & Meinel, C. (2009). Design Thinking – Zur Entwicklung eines explorativen Forschungsansatzes zu einem überprofessionellen Modell. *Neuwerk, Zeitschrift für Designwissenschaft, 1*, 47–61.

McKim, R. H. (1972). *Experiences in visual thinking*. Belmont, CA: Wadsworth Publishing.

McKim, R. H. (2016). Designing for the whole man. In W. J. Clancey (Ed.), *Creative engineering: Promoting innovation by thinking differently* (pp. 198–217). Stanford Digital Repository. Available from http://purl.stanford.edu/jb100vs5745 (Original manuscript 1959).

Meinel, C., & Leifer, L. (2011). Design thinking research. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking. Understand – improve – apply* (pp. xiii–xxxi). Heidelberg: Springer.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. *Proceedings of the ACM CHI 94 human factors in computing systems conference* (pp. 152–158).

Plattner, H., Meinel, C., & Weinberg, U. (2009). *Design thinking*. München: MI Wirtschaftsbuch.

von Thienen, J. P. A., & Meinel, C. (2014). *Let's bring home even more benefits from design thinking: Ideas for an iterated design thinking process model*. Electronic colloquium on design thinking research. http://ecdtr.hpi-web.de/report/2014/001

von Thienen, J. P. A., & Meinel, C. (2015). *Building on a Stages of Change Model to bring home more big design thinking ideas*. Electronic colloquium on design thinking research. http://ecdtr.hpi-web.de/report/2015/001

von Thienen, J. P. A., Clancey, W. J., Corazza, G. E., & Meinel, C. (2017). Theoretical foundations of design thinking. Part I: John E. Arnold's creative thinking theories. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research. Making distinctions: Collaboration versus cooperation* (pp. 13–40). Cham: Springer.

Wilden, A. (1987). *The rules are no game: The strategy of communication*. London: Routledge and Kegan Paul.

Part I
Understanding Success Factors of Design Thinking

Emotions Along the Design Thinking Process

Benedikt Ewald, Axel Menning, Claudia Nicolai, and Ulrich Weinberg

Abstract The creative interaction of a team is where most of the innovation work in organizations happens nowadays. Yet the creative team is an exceptionally messy place in regard to socio-emotional interactions. Working creatively means constantly navigating and negotiating uncertainty and ambiguity, which, apart from constant procedural adaptations, both evokes and needs adequate responses on the socio-emotional level. In this chapter we want to introduce the most important socio-emotional factors for creative teamwork and how the emotional dynamics of the team is shaped by the different phases of the design thinking process. To this end we review automated text analysis of design thinking team meetings as a method to unobtrusively track emotional dynamics throughout the whole design process.

1 Introduction

Creative work in a team nowadays is regarded as the "key building block for innovation and entrepreneurship" (Gilson et al. 2015). We regard a work as creative if its aim is to come up with new and applicable solutions to ambiguous and problems, e.g. coming up with new business models as well as finding a solution to a city planning problem. As much as execution power is needed to succeed with these solutions and eventually end up as an innovation (innovation = idea + implementation)—without the prerequisite of creativity the initial idea will look pretty bleak and lead merely to incremental improvements, if any at all.[1] Despite an abundance of structured innovation models from management theory as well as creativity definitions and measurements from the (cognitive) psychological

[1]We follow Amabile's canonical innovation model here (1996), establishing creativity as "a necessary but not sufficient condition" for innovation, as well as the clear-cut distinction between studying workplace innovation and individual and group creativity as presented by Anderson et al. (2004), based on West and Farr (1990).

B. Ewald (✉) · A. Menning · C. Nicolai · U. Weinberg
Hasso Plattner Institute for Digital Engineering, Potsdam, Germany
e-mail: benedikt.ewald@hpi.de

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_3

perspective available, the phenomenon of creative invention in organizational settings remains elusive and multifaceted. This elusiveness is due to two intertwined aspects. For one, creative work in an organization is influenced by all levels of the system—the individual level, the organizational context as well as the society it is embedded in (Csikszentmihalyi 1999). All these levels come together not only for the production of novel and useful ideas, but also for their evaluation. This makes it hard to pin down isolated sources and factors that linearly lead to a successful innovation. Second, creative work is heavily linked to socio-emotional factors that are not reflected in the traditional economic and managerial models of innovation. The team is, however, a place where the impossibility of definitely attributing an innovation to one actor or factor comes together with the importance of socio-emotional factors for both the creation and acceptance of new ideas in an especially tightly interwoven way. The creative team is a place of heavily interdependent and fast-paced interactions on all communicative levels—procedurally, problem-oriented and socio-emotional. These interactions become particularly complex to chaotic for high performance teams (Losada 1999).

In this chapter we want to point out some of the most important threads that make up this tightly interwoven socio-emotional fabric of a creative team, as well as give an example from our own research of a setup how to study these threads without interfering with or even destroying the delicate fabric of in situ teamwork. The first part of this chapter will present the most important socio-emotional process factors for creative teamwork. The second will illustrate these theoretical underpinnings with a concrete example of how the design thinking process shapes the emotional dynamics of a team, analyzed with the help of the unobtrusive and automatable Linguistic Inquiry and Word Count tool (LIWC).

2 Emotions in (Creative) Teams

In general, moods and emotions play a crucial role in how we approach tasks individually and in a team (Fisher and Ashkanasy 2000). They are considered to be the main mediating mechanism of how work environmental features impact us and how they facilitate teamwork. Especially innovation and creativity are strongly connected to emotions. Emotions accompany and form the whole innovation process, both on the creating and the evaluating and adopting side (Amabile et al. 2005; Gelbrich 2007).

This is due to several factors:

- Innovation and creativity involve navigating both uncertainty and ambiguity due to the nature of the problems treated. In teams, this inherent indeterminism can lead to variant, but equally sound, opinions about how to proceed, leading to disagreements about content and process.

- To be able to come up with and follow through with new, crazy ideas and solutions there has to be a certain psychological safety and support for innovation present at the workplace and in and around the team.
- The individual emotional reactions to uncertainty and goal attainment.

Problems in innovation and design are in general ill-defined—they are so called "wicked problems" (Reitman 1964; Rittel and Webber 1973; Goel and Pirolli 1992). These problems are open, dynamic, complex and networked (Buchanan 1992; Dorst 2015). Due to the nature of the beast, there is no such thing as a right or wrong solution to these problems, but only good-fits or bad-fits, which lead to ambiguous decision-making situations. This ambiguity and the resulting uncertainty involved in both problem and solution finding are not something that needs to be overcome, but is instead "essential to design process" (Bucciarelli 1994, p. 178) and an important driver of creative cognitive processes. Intentionally producing vague expressions is a crucial interactional strategy in design conversations to achieve an effective open and inspirational process (Glock 2009; Zenasni et al. 2008; Christensen and Ball 2017). Being able to deal with ambiguity has also been considered an integral element of a creative personality (Tegano 1990; Tracey and Hutchinson 2016).

Dealing with ambiguity and uncertainty is hard enough alone, but in a team this difficulty increases exponentially. Interpersonal communication per se is rarely unequivocal, and using it to try to reach a common understanding about an ill-defined object can lead to various conflicts and misunderstandings. The relationship between team conflicts and team performance has been explored extensively (see e.g. review by O'Neill et al. 2013). Going back to Jehn's qualitative analysis of 1997, one can distinguish three different types of team conflict: task, relationship, and process conflict.

While the distinction between *task conflicts*—content-related disagreements in the team—and *relationship conflicts*—interpersonal tensions or resentments based on the view of the other person—had been studied before both theoretically and empirically, the notion of *process conflict* was new at this time. Jehn frames a process conflict as "conflict about how task accomplishment should proceed in the work unit, who's responsible for what, and how things should be delegated" (1997, p. 540)—so *how* the team should go about solving the task at hand in contrast to *what* to do or decide on.

The research picture of the latter is relatively clear—both relationship and process conflict have a negative impact on team performance, team satisfaction and team cooperation (de Dreu and Weingart 2003; De Wit et al. 2012; Maltarich et al. 2016).

However, the research on task conflicts is indecisive. It has yielded everything from positive to negative to no relations between task conflicts and team performance (De Wit et al. 2013). This is because the picture for task conflict is complex and relies heavily on other factors and moderators. The relationship between task conflict and team creativity, for example, is mostly "a question of how much and when" (Farh et al. 2010). Early and moderate conflict can be stimulating, while late or too much conflict has no or a detrimental effect (also see Kratzer et al. 2006). If and how conflict can be dealt with is also based on the conflict management style of the team and how it develops through the teamwork (Maltarich et al. 2016). How much task conflicts influence team performance also depends on the complexity of the work (innovation challenge), as it can take away cognitive resources from the task at hand and therefore have a stronger negative impact on complex tasks (de Dreu and Weingart 2003). The biggest threat to team performance coming from task conflict is the co-occurrence of task and relationship conflict (i.e. when a disagreement on the content level becomes intertwined with an interpersonal tension). This can happen coincidentally or causally and leads to more rigidity in decision-making and biased used of information. It negatively affects the performance of the team (de Wit et al. 2013). Task conflict can also turn personal and therefore negatively impact relationship conflict (Guenter et al. 2016). What impact conflict on both team and individual has also depends on the cognitive style of the individual team members (Kim et al. 2012).

How the team deals with conflicts is connected to another prominent socio-emotional factor for innovation: participative safety. Participative safety has been framed as “a shared belief that the team is safe for interpersonal risk taking” (Edmondson 1999, p. 354), which means that the team atmosphere is open and trusting and encourages participation and “speaking up” (West 1990; Peltokorpi and Hasu 2014). Team innovation and creativity needs risk-taking and the courage to pursue crazy ideas as an individual, in a team and on an organizational level (Dewett 2007; Kuczmarski 1996; Edmondson et al. 2001). Participative safety has therefore been considered as a decisive socio-emotional factor for innovation and team performance (Edmondson 1999; Anderson and West 1996). Indeed, it could be said that participative safety mediates the positive effect of task conflict for team performance in general (Bradley et al. 2012). Paradoxically, participative safety has also been shown to be only minimally correlated with innovative outcomes in general (Hülsheger et al. 2009). This is thought to be due to a certain kind of “comfort zone effect” leading to complacency discouraging criticism and divergence for high levels of participative safety. Participative safety could be actually detrimental to creativity. To examine this, Fairchild and Hunter (2014) looked at the combined influence of participative safety and task conflict in the two individual dimensions of creativity, originality and usefulness. They found, as expected, a positive effect of high task conflict on the originality dimensions in the presence of a high participative level. Surprisingly though, the highest positive effect was found in teams low on task conflict *and* participative safety. This points towards the findings of Feist (1998), which reveal that some highly creative individuals score high on openness, but are less conscientious, leading them to prefer a working style with less task interdependence. This could lead to a disconnect among team members, reflected in less attachment and therefore less conflict, but also less participative safety.

But participative safety has also been shown to be positively correlated to learning behavior as well as the quality of being outspoken towards leaders or supervisors (Edmondson 1999; Walumbwa and Schaubroeck 2009). This factor can therefore also lead to a strengthened feeling of confidence of the team itself towards supervisors as well as organizational challenges, which—apart from the core creative work—could facilitate challenges further down the path towards innovation, namely in the implementation phase.

Variability of personalities and backgrounds in a team can also become important when dealing with uncertainty and team goals. Different people respond emotionally different to goal attainment, mediated by the strength of their regulatory focus (Higgins et al. 1997). Regulatory focus theory distinguishes between a promotion- and a prevention-focused approach to life, based on the basic need to attain pleasure and avoid pain (hedonic principle). While the prevention focus is driven by the fear of loss and a certain convention dependence, the promotion focus approaches challenges from the "What do I gain from this?"—perspective. Both are part of the motivational system and facilitate and sustain activation towards a goal (Baas et al. 2011). These foci as well as the activation can be momentary as well as chronic. Activation is a state characterized by increased cognitive, emotional and physiological parameters such as alertness, attentiveness or heart rate (Watson et al. 1999; Brehm and Self 1989). Moderate levels of activation have been shown to enhance creativity (Byron et al. 2010). Before it was assumed that people with a predominantly promotion-oriented focus generally perform better on divergent thinking tasks, just as creating new insights, while prevention focus seems to be connected to analytical problem solving (Crowe and Higgins 1997; Friedman and Forster 2001). Based on Byron et al., it could be shown that this effect is indeed mediated by the level of activation and energy of the creative individual (Baas et al. 2011). Therefore, prevention focus can also promote creativity if the initial goal is not reached, which leads to the activating emotions of anger and frustration. This insight is in line with the findings of the meta-analysis by Baas et al. (2008), which found that activating emotions like happiness and anger are most positively and comprehensively related to creativity. While anxiety has a negative effect on flexibility, deactivating moods like sadness do not have an effect on creative performance at all.

All three aforementioned socio-emotional aspects—conflicts, participative safety and regulatory focus—share a strong reciprocity between each other and the emotional state (e.g., activated/deactivated, positive/anxious) of the creative individual or team (Shin 2014). One promising way of understanding, assessing and improving creative processes is therefore to look at the affective processes needed for and instigated by these (Amabile et al. 2005; Fong 2012). Affect in this context stands for both short-term emotions and more diffusively structured moods (Barsade and Gibson 2007). A group's affective state is a combination of the group's affective composition (bottom-up) and the affective context in which the group is behaving (top-down; Barsade and Gibson 1998). As we are especially interested in the team affect as it is mediated by a specific innovation process, we focus on the bottom-up dynamics of an existent and static team, which excludes organizational context, initial team composition and self-selection effects (see Paulsen and Kauffeld 2016). The emotions expressed by individual teammembers are one starting point for bottom-up dynamics. These can lead to a mood contagion in other team members (Bono and Ilies 2006). It has been shown that positive emotional contagion in teams positively influences work-related outcomes such as improved cooperation, decreased conflict, and increased perception of task performance (Barsade 2002).

The effect of positive affect on creativity is mediated via cognitive and motivational processes (Bjørnebekk 2008; James et al. 2004; Nikitin and Freund 2010). These processes are e.g. neural arousal, divergent thinking and activation of other creativity-generating mechanisms (James et al. 2004). On a team level, a positive mood proved to be beneficial for idea generation and overall creative output by fostering the sharing of ideas (Grawitch et al. 2003; Rhee 2006). A negative tone, in contrast, can undermine an innovation team's reputation and lower a team's performance (Barsade and Knight 2015; Peralta et al. 2015).

2.1 Taking a Process Perspective

Although the mediating mechanisms of team affect on creativity are not fully understood, there is ample evidence that positive affect is beneficial for creativity, both on an individual and a team level. The ensuing question therefore is—how can this positive team effect of affect be achieved? Many consulting approaches to teamwork focus on putting together the right people to achieve successful teamwork. These approaches mostly rely on various personality tests, among which the Myers-Briggs Type Indicator (MBTI) and tests based on the Big Five personality trait model are the most popular. Although some personality traits have been scientifically shown to be beneficial for creative work (primarily openness to experience; McCrae 1987), there are two rather big limitations to this approach. First, apart from the canonical Big Five (e.g. Hahn et al. 2012), most of these tests lack validity and reliability, both internally due to theoretical problems (Boag 2015) and externally in their application for recruitment (Guion and Gottier 1965). Second, even if equipped with valid and reliable results from a personality test these do not carry very far in the heavily interactional and dynamic environment of teamwork. For example, it is still unclear how to determine a "perfect mix" for a team in the first place, as individual characteristics related to creativity say little or nothing at all about their effectiveness at the workplace and in team situations (Feist 1998; Madjar 2008). Accordingly, studies find that "different team compositional mixes will be more or less salient at different periods of performance episodes or stages of team development" (Mathieu et al. 2014, p. 130).

Learning and personal development are key ingredients of today's working world, which conflicts with the static notion of personality traits. Especially creative methodologies and frameworks, such as design thinking, rely on the idea "that individuals and teams have the ability to build their innovative capacity through various tools and methods no matter their predispositions to creativity and innovation" (Martelaro et al. 2015, p. 41). Indeed, it has been shown through extensive meta-analysis that process factors "display stronger links with innovation than input variables" (Hülsheger et al. 2009, p. 1139). In short, and following the canonical "input-process-output" (IPO; e.g. Hackman 1987) model for teamwork one can therefore state: process beats input in all things team.

3 Measuring Emotions Along the Design Thinking Process

To illustrate how a measurement of socio-emotional process factors in vivo can look like, we will dive deeper into a concrete case example: the measurement of emotions expressed in a team along the different phases of the design thinking process.

It has already been shown that a positive team mood also leads to better creative output during ideation (Baas et al. 2008). There also already exists a study looking into the predominance of different regulatory foci in the different design thinking phases by Kröper et al. (2010). However, it assessed only the regulatory focus level and via self-reported Experience Sampling, so not directly examining the emotions uttered.

Here, we will introduce text analysis as an unobtrusive tool to study team affect, present the most important features of the design thinking process and conduct a first analysis of a design thinking team during the key activities of the phases.

3.1 Text Analysis as an Unobtrusive Way to Examine Team Affect

Using text analysis to study team affect presents a new application that has not yet been deeply investigated. This is especially interesting as it presents a way to access the process in an unobtrusive manner. Prior research that focused on process evaluation mainly relied on ratings that are based on videos (e.g., Kauffeld and Lehmann-Willenbrock 2012; Schermuly and Scholl 2012). However, these coding and rating approaches are very time-consuming (e.g., coders using the TEMPO Instrument need nine hours to code one hour of team interaction; Futoran et al. 1989) and would compound the already time-consuming transcription work. Moreover, to ensure adequate reliability it is recommended that at least twenty percent of the video material be coded by two raters to enhance the effort. Nevertheless, error-free analysis is still not ensured. With the promise offered by enhancements in speech-to-text technology in regard to transcription, coding will soon be the rate-limiting step for such analyses. To overcome these problems and eventually move towards near real-time evaluation, we need an approach that is computerized, both to avoid humans as a source of biases and to speed up the evaluation process.

One way to computationally evaluate large amounts of text is the Linguistic Inquiry and Word Count software (LIWC; Pennebaker et al. 2015c). LIWC categorizes words in specific categories and subcategories which can be linked to several psychological models and processes. For example, pronoun use to attentional focus, verb form and word count to social hierarchy and many more (Tausczik and Pennebaker 2010). In its current version, for example, the complete English dictionary contains 6400 entries which are all assigned to one or more of 55 nonexclusive categories, organized under four main themes: basic linguistic processes, psychological processes, personal concerns, and spoken categories (Pennebaker et al. 2015a, b). The LIWC word categories are well validated and their applicability to different contexts of real life interaction has been demonstrated in hundreds of studies (Tauscik and Pennebaker 2010). Research also shows that LIWC is an accurate and valid method to measure emotions in language use, being both in accordance with human ratings (Alpers et al. 2005) and having predictive power (Kahn et al. 2007).

LIWC has been frequently used for sentiment analysis, which analyzes the affectional relationship of individuals towards products, political opinions etc. on a bipolar scale that is usually positive–negative (Stieglitz and Dang-Xuan 2011; Tumasjan et al. 2010). The same approach has also been used to analyze mental health signals in Twitter data (Coppersmith et al. 2014). Although we do not generally look at sentiment (directed affect) but team affect, LIWC has proven to be very effective for such emotional, bipolar evaluations in general (Kahn et al. 2007).

3.2 The Phases of the Design Thinking Process

The conceptual structure of design (thinking) processes can be well approximated along two pairs of concepts: the problem and solution space, and divergent and convergent thinking (Lindberg et al. 2010). The problem-solution pair points to the specific structure of the type of problems which can and can best be solved with a 'designerly' approach—the so-called 'wicked problems' (see above). This special form of complexity mostly stems from the simple fact that these problem fields are inherently *human* problem fields—fields, where bona fide needs and wishes of different users and stakeholders clash, emerge and change continuously ("A design problem keeps changing while it is treated"; Rittel 1987, p. 2). In its approach to take this inherent humanity serious, design and design thinking takes a human-centered, empathetic approach. This approach relies heavily on empathy and a certain form of iterative framing and reframing of the problem. This eventually leads to a co-evolution of problem and solutions space, rather than just a linear development of one predetermined solution (Maher et al. 1996; Dorst and Cross 2001).

A team creative process for innovation generally comprises four stages: identifying the problem; gathering and synthesizing information; generating ideas; and evaluating, refining and selecting them (Gilson et al. 2015). Following Guilford (1950), these four stages can be roughly categorized into the two thinking modes they require—divergent and convergent thinking. Divergent thinking means 'going broad', looking for a high quantity of different insights or ideas with a high semantic diversity (Guilford 1956; Shroyer et al. 2017). The ability of divergent thinking is an important factor for ideational productivity and has been identified with cognitive creativity for decades (Runco and Acar 2012). The underlying assumption of any divergent activity is that a greater number of ideas will lead to more creative outcome. The Torrance test, still the most prevalent creativity test, essentially measures divergent thinking ability (Baer 2011). It has also been argued, that in various framings, this thinking ability of 'going broad' is a *condicio sine qua non* for creativity and innovation processes (e.g., abductive thinking—cf. Dorst 2015; Endrejat and Kauffeld 2017; lateral thinking—De Bono 1968; associative thinking—Mednick 1962). Brainstorming, for example, can be understood as a systematized approach to facilitate divergent thinking (Kalargiros and Manning 2015).

Convergent thinking, in contrast, is needed to boil down the variety of concepts in order to be able to focus and consolidate the explorative work. It consists of the naming, (re-)ordering and linking of concepts. Without convergent thinking, neither the concrete problem treated could be framed nor a final result produced (Cropley 2006; Guilford 1956). Convergent thinking therefore ensures the usefulness of the ideas generated by divergent thinking and prevents a purely novelty-focused "pseudocreativity" (Cattell and Butcher 1968, p. 271).

At the HPI School of Design Thinking these different thinking modes and space explorations are distributed over six phases. These phases are visited repeatedly and iteratively over the course of a full design thinking project, with 'Understand', 'Observe', and 'Synthesis' as part of the problem finding, and 'Ideation', 'Prototype', and 'Test' as part of the solution space. 'Understand', 'Observe' and the 'Ideation' phase can be roughly classified as divergent phases, i.e. as much data or ideas as possible should be gathered and generated, while judgement and selection are deferred. Roughly, because also during 'Understand' a first clustering or ordering can happen, and also the classical brainstorming phase 'Ideation' normally includes a selection of ideas. 'Synthesis' and 'Prototype', in contrast, are convergent phases, i.e. one to a handful of insights or ideas need to be selected, facilitated by certain methods to cluster and filter and make sense out of the vast variety generated before. The distilled insights will then be used to define a 'Point of View' (the pivotal point between problem and solution space), and the selected idea(s) to build a prototype (see Fig. 1). 'Test' can go both ways. This, on the one hand, depends on whether the activity is planning, conducting, or evaluating the testing, and on the other hand, what the outcome of the testing is. While the testing itself is rather divergent in the sense of again going into the field and collecting data in the form of feedback and observations about the interaction of users with the prototype, both planning and especially evaluating the testing can lead to both divergent and convergent thinking modes. Planning, because you can (a) either brainstorm on testing scenarios (divergent), (b) still further refine the prototype and its presentation specifically for testing (convergent), or (c) organize the testing itself (rather convergent or solely procedural). What happens during or right after the evaluation of the testing very much depends on the outcome of the testing. Did the general idea (the critical function) work and resonate with the user? Or does the whole concept need reworking in the form of better understanding the needs of the user (back to research) or coming up with better solutions (back to ideation)? The former would mean to go further towards refinement of the prototype and thereby continuing down the path of convergence. The latter though means collecting (completely) new insights or ideas or at least revisit the ones gathered before, i.e. opening up the solution space again.

Each of the aforementioned phases therefore has its own objectives and key activities, resulting in a different working mode and differently structured outcomes. We capture these key activities and episodes in the form of video and audio recordings of the most decisive moments for the team in terms of content development, i.e. the key team episodes for the respective day. Although the working day for the teams at the HPI School of Design Thinking usually has a preset agenda, the teams diverge from it if needed. This especially happens in the later phases of a full design thinking project, when the first iterations begin.

3.3 Mapping the Emotional Journey of a Design Thinking Team

The different objectives and ensuing working modes of the different phases effect the emotional journey of the design thinking team differently, along with the general development of the team's climate over time. We want to illustrate how studying affect via text analysis can help to better understand the effect of both the general structure of the design thinking phases as well as, in the future, of the specific methods employed to reach the phase objective. To this end we will map out the emotional journey of one team working on a 6-weeks design thinking project in the dimensions of positive versus negative affect in the following part.

3.3.1 The Data

The team recorded worked on a real-life design thinking project with project partners from the music industry for two full days per week, over the course of 8 weeks. All in-house project (meeting and working) days were audio- and videotaped to capture the most decisive phases of the design thinking process. These phases were: (1) first approximation of the challenge ('Understand'), (2) bringing together and synthesizing information gathered in the field ('Synthesis'; 'Point of View' in Fig. 1), (3) developing ideas ('Ideation'), (4) preparing the 'Testing' and (5) building and refining prototypes ('Prototype'). The two phases 'Observe' and the testing itself are conducted outside of a traditional team meeting setting in the field and as such are not included in our analyses. The video and audio data were transcribed and segmented into speaker turns. A turn begins when a new speaker starts her utterance and ends when she stops to speak or is interrupted.

The transcribed text was analyzed with the LIWC2015 software. The LIWC2015 Dictionary comprises almost 6400 words, word stems, and selected emoticons, all of which are categorized into one or more categories and subcategories. For our

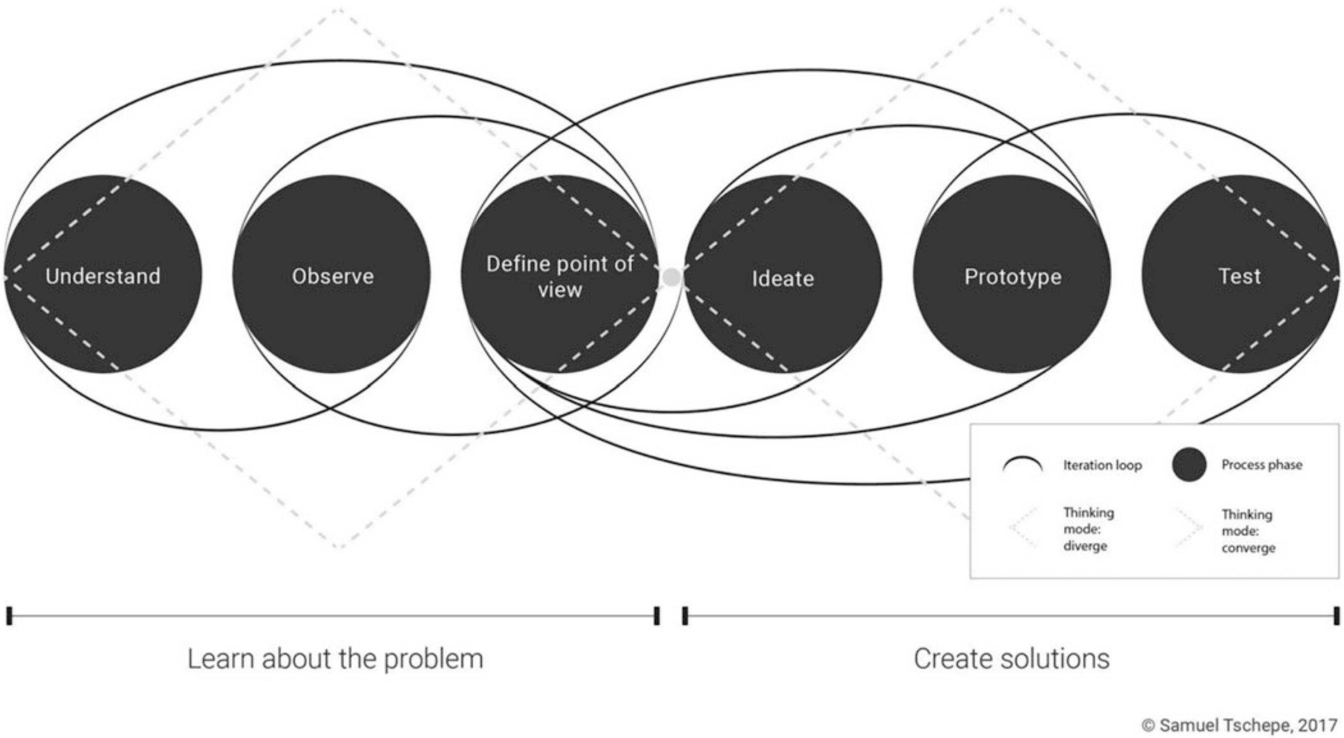


Fig. 1 The design process used at the HPI School of Design Thinking with the six phases ‘Understand’, ‘Observe’, ‘Define point of view’, ‘Ideate’, ‘Prototype’ and ‘Test’; with an overlay indicating the convergent and divergent phases (dotted line) as well as indicating the problem (“Learn about the problem”) and the solution phases (“Create solutions”; Tschepe 2017)

Table 1 LIWC natural speech means vs the LIWC means of the design thinking team

	Phase	Turns	WC	Affect	Posemo	Negemo
LIWC	Natural speech means		–	6.5	5.3	1.2
Design thinking team	Total/Average	1570	21,228	13.0	11.7	1.3
	Understand	321	4948	13.4	12.2	1.0
	Synthesis	265	4140	9.3	8.0	1.2
	Ideation	267	4633	12.9	11.5	1.3
	Testing Feedback	355	4656	16.1	14.4	1.6
	(Refining) Prototype	362	2851	13.4	12.1	1.2

analysis, we make use of the *affect* categories *posemo* (i.e. including positive emotion words) and *negemo* (i.e. including negative emotion words).

The *affect* category comprises 1393 words in total, reflecting different affective states. The two mutually exclusive and mutually exhaustive first order subcategories are *posemo* with 620 words reflecting positive emotions (e.g. 'love', 'sweet', 'nice') and *negemo* with 744 words reflecting negative emotions (e.g. 'hurt', 'ugly', 'nasty'; Pennebaker et al. 2015c).

3.3.2 Communicative Style in General

We included 1570 statements (21,228 words) uttered by the design team during the key episodes of the above-mentioned activities in our analysis.

With an overall value of 13.0 the *affect* average of the design thinking (DT) team is much higher than the means for natural speech provided by the LIWC Development manual of 6.5 (Pennebaker et al. 2015c). This above-average affectivity is predominantly due to positive emotions voiced (*posemo*; LIWC2015 natural speech mean = 5.3, DT team = 11.7), whereas the expression of negative emotions (*negemo*) stays on an average level (see Table 1).

The design thinking team therefore seems to be generally more emotional, and of these emotions expressed above-average are positive.

3.3.3 Positive and Negative Emotions in the Different Phases

In Fig. 2 we can see graphically, what the numbers in Table 1 describe—after a very positive start in the 'Understand' phase, expression of positive emotions (*posemo*) drops about to about three quarter of the initial value in the 'Synthesis' phase. It then again rises during 'Ideation', until it reaches its high point when collecting the feedback of the 'Testing'. *Posemo* then again levels out at nearly the same value during refining the 'Prototype' (12.2 to 12.1) as it was during the first phase.

The negative emotion category, *negemo*, in contrast shows no notable variation and stays very close to the general LIWC2015 *negemo* mean for natural speech of 1.2 (see Table 1).

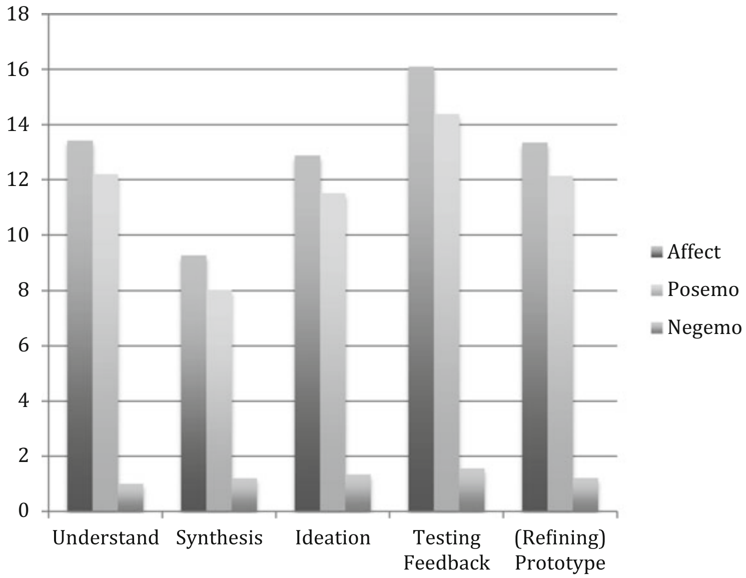


Fig. 2 The *affect*, *posemo* and *negemo* averages of the investigated design thinking team along the key activities

This result indicates that—like already in the overall mean—also nearly all of the affect dynamics are due to the dynamics of positive emotions expressed.

Especially noteworthy is the sharp drop in the ‘Synthesis’ phase. It is here where the problem and the solution space explicitly get connected via the ‘Point of View’. Distilling the “best” insights from an often verwhelming mass of data collected during ‘Observation’ is also often the hardest part, especially for less experienced design thinkers. But this difficulty also is attributable to a decisively different working mode necessary during ‘Synthesis’—in contrast to the propagated and prototypical diverging thinking mode in design thinking. ‘Synthesis’ needs clustering, selecting and—hence the name—synthesizing information, rather than going broad and crazy. This convergent thinking mode can come with some serious challenges to design thinking teams. The comparison, discussion and subsequent selection of insights puts the team under argumentative and judgmental pressure, potentially resulting in more conflicts and also a higher cognitive load (Kim et al. 2012). Studies on transactive memory have shown that especially under time pressure teams with a similar level of expertise about the problem at hand suffer from a higher cognitive load (Hollingshead 1998). In general, it has been shown that the process of weighing arguments comes with a higher mental effort than constructing new solutions (Shehab and Nussbaum 2015). A greater cognitive load finally can lead to a reduction in emotional intensity (Van Dillen and Koole 2007; Van Dillen et al. 2009).

The importance of the convergent movement has been highlighted before, and the 'Synthesis' can often be a decisive working phase for a design thinking project. It is therefore no accident that experienced designers are said to be especially good 'synthesizers' (Goldschmidt and Tatsa 2005).

Maybe it is this cognitive difficulty what we can see in the data here, leading to a drop in (positive) emotionality. But to prove this hypothesis more than this first probe will be needed and should be subject to further, more comprehensive studies.

The global maximum we see for the collection of the 'Testing' feedback can in contrast be quite easily explained when we look into the real-life data. The feedback for the prototype of the team was just very positive, which was reflected in a positive team mood. This as well could be an interesting starting point for further studies, especially in combination with accounting for the different cognitive style or regulatory foci of the team members, as this will influence the reaction to positive or negative feedback.

For the end of this chapter though this first illustration of how LIWC can be used to probe the emotional journey of a design thinking team shall be sufficient.

4 Conclusion

Our pre-study shows how LIWC can be used as an unobtrusive instrument to examine the affect dynamics of an innovation team in the wild. Affect analysis is an important step towards studying other, more complex socio-emotional constructs introduced in Sect. 2, like regulatory focus, intragroup conflict, and participative safety on a process-level. It has already been shown that the closer a project approaches its end, the more important a team's affect for outcomes becomes (Paulsen et al. 2016). Thus, future research following the phase structure of innovation processes might also look at the different impact and relevance of team affect in the different phases. Also, the interplay between different constructs as well as the manageability of these is an important field to further look into, especially for real-life application in education and industry. Such as participative safety can mitigate high conflict levels, which in turn can be due to diverging momentary regulatory foci, there are several other mechanisms and interdependencies which need to be better understood. For example, there are already studies on how to "hack" into appraisal processes to improve creative ideation (De Rooij et al. 2015), as these are also connected to emotional reactions via regulatory focus (see above). Another example is that loading the working memory of individuals can reduce negative mood and make them more robust against negative stimuli (Van Dillen and Koole 2007). To make these and more insights frugal for creative team work would also mean to look deeper into single activities instead of broad phases to evaluate the methods being used in regard to how they affect the emotional level of the teams. This could be used to formulate recommendations to innovation practitioners and facilitators alike on which methods to employ and when. Such recommendations would be particularly helpful for team and leadership coaches in becoming aware of the importance of coaching interventions on the emotional level. This would allow them, as well as (future) members of creative teams in general, to develop the emotional intelligence much needed as enablers of creativity and innovation (see Ashkanasy and Dasborough 2003; Ashkanasy and Daus 2005; Goleman and Boyatziz 2008).

The study of groups over time in general is still underrepresented (Cronin and Weingart 2011), especially in the wild. Laboratory settings and tasks are often not engaging enough to measure realistic emotional involvement (Shin 2014; George and King 2007). Text analysis makes both an unobtrusive and high-resolution process study of team dynamics over time and in the wild possible, as it only needs (though proper quality) audio data. Text analysis also does not need any coder-training, tests for inter-rater reliability or the high investment of time for coding itself. The transcription effort though still remains.

Although LIWC has its limitations due its lexical and therefore static nature, as newly designed categories need to be validated anew individually. The next step here lies with more adaptable machine-learning algorithms (e.g., Empath by Fast et al. 2016). Such computational tools, lastly, also open up many more opportunities for statistical analysis, which may be more revealing for implicit emotional processes than a qualitative analysis of language (Paletz et al. 2017).

For the concrete study of team affect through the different phases, this case study of course only has very limited validity. We are currently working on making more text data of teams and phases available. Especially the difference between divergent and convergence phases is something to look into more in detail and quantitatively. Additional measures, such as self-reports and outcome measures, could be included and would allow for a comparison between reported and observed emotions, as well as the collection of data on other, related socio-emotional constructs. Also, the outcome should be taken into consideration to be able to be able to identify how the emotional journey of a successful team differs from an unsuccessful team.

References

Alpers, G. W., Winzelberg, A. J., Classen, C., Roberts, H., Dev, P., Koopman, C., & Taylor, C. B. (2005). Evaluation of computerized text analysis in an Internet breast cancer support group. *Computers in Human Behavior, 21*(2), 361–376.

Amabile, T. M. (1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, CO: Westview Press.

Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative Science Quarterly, 50*, 367–403.

Anderson, N., & West, M. A. (1996). The team climate inventory: Development of the TCI and its applications in teambuilding for innovativeness. *European Journal of Work and Organizational Psychology, 5*(1), 53–66.

Anderson, N., De Dreu, C. K. W., & Nijstad, B. A. (2004). The routinization of innovation research: A constructively critical review of the state-of-the-science. *Journal of Organizational Behavior, 25*, 147–173.

Ashkanasy, N. M., & Dasborough, M. T. (2003). Emotional awareness and emotional intelligence in leadership teaching. *Journal of Education for Business, 79*(1), 18–22.

Ashkanasy, N. M., & Daus, C. S. (2005). Rumors of the death of emotional intelligence in organizational behavior are vastly exaggerated. *Journal of Organizational Behavior, 26*(4), 441–452.

Baas, M., De Dreu, C. K. W., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood – creativity research: Hedonic tone, activation, or regulatory focus? *Psychological Bulletin, 134* (6), 779–806.

Baas, M., De Dreu, C. K. W., & Nijstad, B. A. (2011). When prevention promotes creativity: The role of mood, regulatory focus, and regulatory closure. *Journal of Personality and Social Psychology, 100*(5), 794–809.

Baer, J. (2011). How divergent thinking tests mislead us: Are the Torrance Tests still relevant in the 21st century? The Division 10 debate. *Psychology of Aesthetics, Creativity, and the Arts, 5*(4), 309–313.

Barsade, S. (2002). The Ripple Effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly, 47*, 644–675.

Barsade, S. G., & Gibson, D. E. (1998). Group emotion: A view from top and bottom. *Research on Managing Groups and Teams, 1*(4), 81–102.

Barsade, S. G., & Gibson, D. E. (2007). Why does affect matter in organizations? *Academy of Management Perspectives, 21*(1), 36–59.

Barsade, S. G., & Knight, A. P. (2015). Group affect. *Annual Review of Organizational Psychology and Organizational Behavior, 21*(2), 21–46.

Bjørnebekk, G. (2008). Positive affect and negative affect as modulators of cognition and motivation: The rediscovery of affect in achievement goal theory. *Scandinavian Journal of Educational Research, 52*, 153–170.

Boag, S. (2015). Personality assessment, "construct validity", and the significance of theory. *Personality and Individual Differences, 84*, 36–44.

Bono, J. E., & Ilies, R. (2006). Charisma, positive emotions and mood contagion. *Leadership Quarterly, 17*(4), 317–334.

Bradley, B. H., Postlethwaite, B. E., Klotz, A. C., Hamdani, M. R., & Brown, K. G. (2012). Reaping the benefits of task conflict in teams: The critical role of team psychological safety climate. *Journal of Applied Psychology, 97*(1), 151–158.

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology, 40*, 109–131.

Bucciarelli, L. L. (1994). *Designing engineers*. Cambridge, MA: MIT Press.

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues, 8*(2), 5–21.

Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: A metaanalysis examining competing theoretical models. *Journal of Applied Psychology, 95*(1), 201.

Cattell, R. B., & Butcher, H. J. (1968). *The prediction of achievement and creativity*. New York: Bobbs-Merrill.

Christensen, B. T., & Ball, L. J. (2017). Fluctuating epistemic uncertainty in a design team as a metacognitive driver for creative cognitive processes. *CoDesign, Special Issue on Designing Across Cultures*, 1–20.

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality* (pp. 51–60).

Cronin, M. a., & Weingart, L. R. (2011). Dynamics in groups: Are we there yet? *Academy of Management Annals, 5*(November), 37–41.

Cropley, A. (2006). In praise of convergent thinking. *Creativity Research Journal, 18*(3), 391–404.

Crowe, E., & Higgins, E. T. (1997). Regulatory focus and strategic inclinations: Promotion and prevention in decision-making. *Organizational Behavior and Human Decision Processes, 69* (2), 117–132.

Csikszentmihalyi, M. (1999). 16 implications of a systems perspective for the study of creativity. In *Handbook of creativity* (pp. 313–335). Cambridge: Cambridge University Press.
De Bono, E. (1968). *New think: The use of lateral thinking in the generation of new ideas*. New York: Avon Books.
De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. *Journal of Applied Psychology, 88*(4), 741–749.
de Rooij, A., Corr, P. J., & Jones, S. (2015). Emotion and creativity: Hacking into cognitive appraisal processes to augment creative ideation. In *SIGCHI conference on creativity and cognition*. Glasgow.
Dewett, T. (2007). Linking intrinsic motivation, risk taking, and employee creativity in an R&D environment. *R&D Management, 37*(3), 197–208.
De Wit, F. R. C., Greer, L. L., & Jehn, K. A. (2012). The paradox of intragroup conflict: A meta-analysis. *Journal of Applied Psychology, 97*(2), 360–390.
De Wit, F. R. C., Jehn, K. A., & Scheepers, D. (2013). Task conflict, information processing, and decision-making: The damaging effect of relationship conflict. *Organizational Behavior and Human Decision Processes, 122*, 177–189.
Dorst, K. (2015). *Frame innovation: Create new thinking by design*. Cambridge: MIT Press.
Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem-solution. *Design Studies, 22*(5), 425–437.
Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350–383.
Edmondson, A. A. C., Bohmer, R. R. M., & Pisano, G. P. (2001). Disrupted routines: Team learning and new technology implementation in hospitals. *Administrative Science Quarterly, 46* (4), 685–716.
Endrejat, P. C., & Kauffeld, S. (2017). Wie könnten wir Organisationsentwicklungen partizipativ gestalten? *Gruppe. Interaktion. Organisation. Zeitschrift Für Angewandte Organisationspsychologie (GIO)*, *48*(2), 143–154.
Fairchild, J., & Hunter, S. T. (2014). "We've got creative differences": The effects of task conflict and participative safety on team creative performance. *Journal of Creative Behavior, 48*(1), 64–87.
Farh, J.-L., Lee, C., & Farh, C. I. C. (2010). Task conflict and team creativity: A question of how much and when. *Journal of Applied Psychology, 95*(6), 1173–1180.
Fast, E., Chen, B., & Bernstein, M. S. (2016). Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI conference on human factors in computing systems – CHI'16* (pp. 4647–4657).
Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review, 2*(4), 290–309.
Fisher, C. D., & Ashkanasy, N. M. (2000). The emerging role of emotions in work life: An introduction. *Journal of Organizational Behavior, 21*(2), 123–129.
Fong, C. T. (2012). The effects of emotional ambivalence on creativity. *The Academy of Management Journal, 49*(5), 1016–1030.
Friedman, R. S., & Forster, J. (2001). The effects of promotion and prevention cues on creativity. *Journal of Personality and Social Psychology, 81*(6), 1001–1013.
Futoran, G. C., Kelly, J. R., & McGrath, J. E. (1989). TEMPO: A time-based system for analysis of group interaction process. *Basic and Applied Social Psychology, 10*(3), 211–232.
Gelbrich, K. (2007). *Innovation und Emotion. Die Funktion von Furcht und Hoffnung im Adoptionsprozess einer technologischen Neuheit für die Kunststoffbranche*. Göttingen: Cuvillier Verlag.
George, J. M., & King, E. B. (2007). Potential pitfalls of affect convergence in teams: Functions and dysfunctions of group affective tone. In E. A. Mannix, M. A. Neale, & C. P. Anderson (Eds.), *Research on managing groups and teams* (Vol. 10, pp. 97–124). Greenwich, CT: JAI Press.

Gilson, L. L., Lim, H. S., Litchfield, R. C., & Gilson, P. W. (2015). Creativity in teams: A key building block for innovation and entrepreneurship. In *The Oxford handbook of creativity, innovation and entrepreneurship* (pp. 177–204). Oxford: Oxford University Press.

Glock, F. (2009). Aspects of language use in design conversation. *CoDesign, 5*(1), 5–19.

Goel, V., & Pirolli, P. (1992). The structure of design problem spaces. *Cognitive Science, 16*(3), 395–429.

Goldschmidt, G., & Tatsa, D. (2005). How good are good ideas? Correlates of design creativity. *Design Studies, 26*(6), 593–611.

Goleman, D., & Boyatziz, R. (2008). Social intelligence and the biology of leadership. *Harvard Business Review, 86*(9), 74–81.

Grawitch, M. J., Munz, D. C., Elliott, E. K., & Mathis, A. (2003). Promoting creativity in temporary problem-solving groups: The effects of positive mood and autonomy in problem definition on idea-generating performance. *Group Dynamics: Theory, Research, and Practice, 7*, 200–213.

Guenter, H., van Emmerik, H., Schreurs, B., Kuypers, T., van Iterson, A., & Notelaers, G. (2016). When task conflict becomes personal. *Small Group Research, 47*(5), 569–604.

Guilford, J. P. (1950). Creativity. *American Psychologist, 5*, 444–454.

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin, 53*, 267–293.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*(2), 135–164.

Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice-Hall.

Hahn, E., Gottschling, J., & Spinath, F. M. (2012). Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory (BFI-S). *Journal of Research in Personality, 46* (3), 355–359.

Higgins, E. T., Shah, J., & Friedman, R. (1997). Emotional responses to goal attainment: Strength of regulatory focus as moderator. *Journal of Personality and Social Psychology, 72*(3), 515–525.

Hollingshead, A. B. (1998). Communication, learning, and retrieval in transactive memory systems. *Journal of Experimental Social Psychology, 34*(5), 423–442.

Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *The Journal of Applied Psychology, 94*(5), 1128–1145.

James, K., Brodersen, M., & Eisenberg, J. (2004). Workplace affect and workplace creativity: A review and preliminary model. *Human Performance, 17*, 169–194.

Jehn, K. A. (1997). A qualitative analysis of conflict types and dimensions in organizational groups. *Administrative Science Quarterly, 42*, 530–557.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology, 120*(2), 263–286.

Kalargiros, E. M., & Manning, M. R. (2015). Divergent thinking and brainstorming in perspective: Implications for organization change and innovation. In A. B. Shani & D. A. Noumair (Eds.), *Research in organizational change and development. Research in organizational change and development* (Vol. 23, pp. 293–327). Bingley: Emerald Group.

Kauffeld, S., & Lehmann-Willenbrock, N. (2012). Meetings matter: Effects of team meetings on team and organizational success. *Small Group Research, 43*(2), 130–158.

Kim, J. M., Choi, J. N., & Park, O. S. (2012). Intuitiveness and creativity in groups: Cross-level interactions between group conflict and individual cognitive styles. *Social Behavior and Personality, 40*(9), 1419–1434.

Kratzer, J., Leenders, R. T. A. J., & van Engelen, J. M. L. (2006). Team polarity and creative performance in innovation teams. *Creativity and Innovation Management, 15*(1), 96–104.

Kröper, M., Fay, D., Lindberg, T., & Meinel, C. (2010). Interrelations between motivation, creativity and emotions in design thinking processes – An empirical study based on regulatory

focus theory. In *Proceedings of the 11th international conference on design creativity (ICDC2010)*, (November) (pp. 97–104).
Kuczmarski, T. D. (1996). What is innovation? The art of welcoming risk. *Journal of Consumer Marketing, 13*(5), 7–11.
Lindberg, T., Noweski, C., & Meinel, C. (2010). Evolving discourses on design thinking: How design cognition inspires meta-disciplinary creative collaboration. *Technoetic Arts: A Journal of Speculative Research, 8*(1), 31–37.
Losada, M. (1999). The complex dynamics of high performance teams. *Mathematical and Computer Modeling, 30*(9), 179–192.
Madjar, N. (2008). Emotional and informational support from different sources and employee creativity. *Journal of Occupational and Organizational Psychology, 81*, 83–100.
Maher, M. L., Poon, J., & Boulanger, S. (1996). Formalising design exploration as co-evolution: A combined gene approach. In *Advances in formal design methods for CAD: Proceedings of the IFIP WG5.2 workshop on formal design methods for computer-aided design* (pp. 3–30).
Maltarich, M. A., Kukenberger, M., Reilly, G., & Mathieu, J. (2016). Conflict in teams: Modeling early and late conflict states and the interactive effects of conflict processes. *Group & Organization Management, 42*(2), 1–32.
Martelaro, N., Ganguly, S., Steinert, M., & Jung, M. (2015). The personal trait myth: A comparative analysis of the innovation impact of design thinking tools and personal traits Nikolas. In *Design thinking research: Understanding innovation* (pp. 41–57).
Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and integration of team composition models: Moving toward a dynamic and temporal framework. *Journal of Management, 40*(1), 130–160.
McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52*(6), 1258–1265.
Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69*(3), 220–232.
Nikitin, J., & Freund, A. M. (2010). When wanting and fearing go together: The effect of co-occurring social approach and avoidance motivation on behavior, affect, and cognition. *European Journal of Social Psychology, 40*, 783–804.
O'Neill, T. A., Allen, N. J., & Hastings, S. E. (2013). Examining the "Pros" and "Cons" of team conflict: A team-level meta-analysis of task, relationship, and process conflict. *Human Performance, 26*(3), 236–260.
Paletz, S. B. F., Chan, J., & Schunn, C. D. (2017). The dynamics of micro-conflicts and uncertainty in successful and unsuccessful design teams. *Design Studies, 50*, 39–69.
Paulsen, H. F. K., & Kauffeld, S. (2016). Ansteckungsprozesse in Gruppen: Die Rolle von geteilten Gefühlen für Gruppenprozesse und -ergebnisse. *Gruppe. Interaktion. Organisation. Zeitschrift Für Angewandte Organisationspsychologie (GIO), 47*(4), 357–364.
Paulsen, H. F. K., Klonek, F. E., Schneider, K., & Kauffeld, S. (2016). Group affective tone and team performance: A week-level study in project teams. *Frontiers in Communication, 1* (November), 1–10.
Peltokorpi, V., & Hasu, M. (2014). How participative safety matters more in team innovation as team size increases. *Journal of Business and Psychology, 29*(1), 37–45.
Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015a). *LIWC2015 – Operator's Manual*.
Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015b). *Linguistic Inquiry and Word Count: LIWC2015*. Pennebaker Conglomerates, Austin, 2015.
Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015c). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin.
Peralta, C. F., Lopes, P. N., Gilson, L. L., Lourenço, P. R., & Pais, L. (2015). Innovation processes and team effectiveness: The role of goal clarity and commitment, and team affective tone. *Journal of Occupational and Organizational Psychology, 88*(1), 80–107.

Reitman, W. R. (1964). Heuristic decision procedures, open constraints, and the structure of ill-defined problems. In *Human judgments and optimality* (pp. 282–315). New York: Wiley.
Rhee, S. Y. (2006). Shared emotions and group effectiveness: The role of broadening-and-building interactions. In K. M. Weaver (Ed.), *Proceedings of the sixty-fifth annual meeting of the academy of management*.
Rittel, H. (1987). The reasoning of designers. *Arbeitspapier zum international congress on planning and design theory*.
Rittel, H., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4* (2), 155–169.
Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66–75.
Schermuly, C. C., & Scholl, W. (2012). The Discussion Coding System (DCS)—A new instrument for analyzing communication processes. *Communication Methods and Measures, 6*(1), 12–40.
Shehab, H. M., & Nussbaum, E. M. (2015). Cognitive load of critical thinking strategies. *Learning and Instruction, 35*, 51–61.
Shin, Y. (2014). Positive group affect and team creativity. *Small Group Research, 45*(3), 337–364.
Shroyer, K., Turns, J., Lovins, T., Cardella, M., & Atman, C. (2017). Team idea generation in the wild: A view from four timescales. In B. T. Christensen, L. J. Ball, & K. Halskov (Eds.), *Analysing design thinking: Studies of cross-cultural co-creation* (pp. 521–540). Leiden: CRC Press.
Stieglitz, S., & Dang-Xuan, L. (2011). The role of sentiment in information propagation on Twitter —An empirical analysis of affective dimensions in political tweets. In *Proceedings of the 22nd Australasian Conference on Information Systems (ACIS), Sydney (Australia)* (Vol. 38). Retrieved from http://aisel.aisnet.org/acis
Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54.
Tegano, D. W. (1990). Relationship of tolerance of ambiguity and playfulness to creativity. *Psychological Reports, 66*(3), 1047–1056.
Tracey, M. W., & Hutchinson, A. (2016). Uncertainty, reflection, and designer identity development. *Design Studies, 42*, 86–109.
Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM, 10*(1), 178–185.
Van Dillen, L. F., & Koole, S. L. (2007). Clearing the mind: A working memory model of distraction from negative mood. *Emotion, 7*, 715–723.
Van Dillen, L. F., Heslenfeld, D. J., & Koole, S. L. (2009). Tuning down the emotional brain: An fMRI study of the effects of cognitive load on the processing of affective images. *Neuroimage, 45*(4), 1212–1219.
Walumbwa, F. O., & Schaubroeck, J. (2009). Leader personality traits and employee voice behavior: Mediating roles of ethical leadership and work group psychological safety. *Journal of Applied Psychology, 94*(5), 1275–1286.
Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*(5), 820–838.
West, M. A. (1990). The social psychology of innovation in groups. In M. A. West & J. L. Farr (Eds.), *Innovation and creativity at work: Psychological and organizational strategies*. Chichester: Wiley.
West, M. A., & Farr, J. L. (1990). Innovation at work. In M. A. West & J. L. Farr (Eds.), *Innovation and creativity at work: Psychological and organizational strategies*. Chichester: Wiley.
Zenasni, F., Besançon, M., & Lubart, T. (2008). Creativity and tolerance of ambiguity: An empirical study. *Journal of Creative Behavior, 42*(1), 61–72.

Measuring Design Thinking Practice in Context

Adam Royalty, Helen Chen, Bernard Roth, and Sheri Sheppard

Abstract Design thinking is often misconceived as a competency that a person or organization either has or does not have. This is problematic because that perspective can lead to an incorrect assumption that design thinking is uniformly applied at the same level by anyone who knows it. This chapter describes design thinking as a practice that can range greatly depending on context. It outlines a series of measures developed to highlight both different aspects of design thinking and how those aspects can vary. These measures provide a more detailed assessment of what is necessary to successfully apply design thinking.

1 Introduction

This chapter has two primary goals: (1) to develop research-based measures that can detect how a variety of design thinking trainings lead to a range of design thinking practice; and (2) continue building on the existing work focused on measuring how design thinking is applied in organizational settings.

This work is important because there is increasing demand from companie and institutions of education around the world for design thinking. In fact, beyond design thinking, the field of design is spreading across businesses at an incredible pace

Bernard Roth and Sheri Sheppard are Principal Investigators (PI) for this chapter.

A. Royalty (✉)
Hasso Plattner Institute of Design (d.school), Stanford, CA, USA
e-mail: adam@dschool.stanford.edu

H. Chen · S. Sheppard
Department of Mechanical Engineering, Stanford, CA, USA
e-mail: hlchen@stanford.edu; sheppard@stanford.edu

B. Roth
Hasso Plattner Institute of Design (d.school), Stanford, CA, USA

Department of Mechanical Engineering, Stanford, CA, USA
e-mail: broth@stanford.edu

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_4

(Kolko 2015) and is influencing organizations' strategy, working culture, and more. As design becomes more integral to workplace culture, the need to measure and assess its influence becomes more critical since organizations will only continue to invest in design thinking as long as they can understand and document its impact in both qualitative and quantitative terms. And while design thinking measures do exist (e.g., Hawthorne et al. 2016; Royalty et al. 2014, 2015; Saggar et al. 2015) they need to be more robust in order to be effective in real world contexts.

To this end, there are two lines of work outlined in this chapter. The first line centers on capturing how individuals and teams progress through a design challenge. For the initial phase, the setting included courses at the Hasso Plattner Institute of Design at Stanford (Stanford d.school). This more controlled setting allowed the research team to establish a suite of measures that could then be used in other companies and non-profits.

The second line of work is the development of design thinking ecologies. This builds on prior work focusing on how design thinking exists within organizations (Royalty and Roth 2016; Royalty and Sheppard 2018). The original impetus behind this work still holds true: organizations need a way to compare and contrast how design thinking is applied in real world settings in order to continue to learn and grow. One aspect that has been added to the design thinking ecology framework is a way to capture how individuals interface with the organizational ecosystem as a whole.

Both lines will contribute to research on design thinking by expanding the number and types of tools researchers can use to measure impact. Furthermore, the measures and activities themselves should have a positive impact on how design thinking is taught and its application in both academic contexts and organizational environments.

2 Measuring a Range of Design Thinking Practice

2.1 Background

One of the primary misunderstandings design thinking works to correct is that some people are creative and some people are not. Much academic work has been done to show that creativity is much more nuanced than a belief that you either have it or you don't (Jablokow and Kirton 2009; Kaufman and Beghetto 2009; Richards 1990; Torrance 1988). However, many people still see themselves as not creative. In fact, that was one of the driving forces behind the creation of the Stanford d.school (Kelley and Kelley 2013). This same false binary extends to design. It is true that there are professional designers, many of whom have skills that are categorically more advanced as compared with people without design training. This does not mean that non-designers cannot practice design. The existence of professional mathematicians does not preclude the ability of everyone else to do math. It is important that design thinking is seen the same way as math.

It is tempting to see oneself as a design thinker after completing design thinking coursework or a workshop. Although these activities help people progress, it does not mean they are necessary prerequisites to become a design thinker. Nor does it mean that all graduates practice at the same level of design thinking ability. This last idea is especially important. If design thinking is viewed as all or nothing, then there is less of an incentive to continue to improve one's skills once someone is perceived to "have it." Furthermore, it may be the case that people whose practice is still relatively weak actually apply design thinking in an incorrect or ineffective way. This can lead to unfair judgments or erroneous conceptions on what design thinking is as a whole. In other words, people might come to believe that design thinking does not have value, when in fact, it simply was not performed well. This can have a negative impact and curtail the entire design thinking movement.

There is evidence that design thinking is not a binary outcome but rather an iterative way of working that people learn over time. There are different levels of understanding on the way to mastery (Beckman and Barry 2007; Rauth et al. 2010). Furthermore, the amount of training demonstrates a measurable impact on design thinking dispositions. People who took intensive trainings—with between 20 and 25 hours of contact time—demonstrated a much greater increase in creative agency than participants who took a less intensive training—10 hours of contact time (Kienitz et al. 2014; Royalty et al. 2014, 2015). The scale that measures creative agency is a useful tool to find an individual outcome of a learning experience. However, it does not focus on how people are actually practicing design thinking, it examines a psychological component. A design thinking practice measure is a needed addition to the existing measurement tools because such a measure will give researchers and practitioners a stronger sense for what people using design thinking are actively doing during their problem solving. This is important because, among other things, it will allow facilitators to assess and give feedback on the actions people take in addition to the dispositions they hold.

Practice is context dependent. In an organizational context, design thinking might be practiced differently depending on the goals of the organization (Royalty and Roth 2016). For example, if a company wants to better understand its customers, design practices like empathy and reframing are going to show up more. As the development of these measures is still in a nascent stage, it makes sense to build them first in a more controlled environment before applying them in a real world setting. That is why the current line of work focuses first on capturing the design thinking practice students demonstrate in the Stanford d.school. This environment is more controlled because the ultimate purpose is one of learning and the work all takes place in courses. Design practice in an academic setting is still very relevant to practice in industry since student teams progress through challenges using the same methods and mindsets as teams in real world settings. Furthermore, the way people develop design thinking capacity in both settings is very similar.

If one accepts that the level of design thinking practice ranges it is important to acknowledge that the level of design thinking instruction practice also ranges. It would be short sighted to measure design thinking practice without considering how it was taught to people in the first place. There are dozens of universities, consultancies, and internal organizational training programs that conduct design thinking trainings (Köppen et al. 2016). However, these courses and workshops are taught by people with varying levels of background in both design and conducting experiential education. There is a great deal of research describing the effect teachers have on student learning and transfer (Ambrose et al. 2010; Bransford and Schwartz 1999; Hattie 2003). Therefore it is worth exploring how instruction impacts the students' practice.

Other variables that impact instruction are structure and intent. The formal teaching of design thinking has grown over the past 10 years. Most of the initial curriculum was based on specific learning experiences developed at the Stanford d. school and the HPI D-School. For example the "wallet" or "gift-giving" project created by the Stanford d.school is a common introduction to design thinking run around the world. There are also myriad "bootcamps" based on workshops pioneered by the Stanford d.school and the HPI D-School. The 3-day workshop became something of an industry standard. Increasingly, the structure of these experiences has begun to evolve. A number of universities offer short courses, design thinking hackathons, week-long workshops, etc. *How does the structure of a learning experience affect practice?*

Instructional intent is also shifting. The Stanford d.school recently reclassified the courses it hosts as *core*, *boost*, and *pop-out* (d.school 2017). The new classification is based in part on the eight design abilities the institute identified as being crucial for the practice of design thinking (d.school 2017). *Core courses* teach a broad overview of design thinking. They tend to cover all eight design abilities. *Boost courses* give students the opportunity to work in depth on three to four design abilities. *Pop-out courses* are very short and focus on just one or two design abilities. Just as is the case with the teacher effect, the structure and intent of learning experiences need to be taken into account when measuring a courses contribution to design thinking practice.

After narrowing in on the importance of practice and the variables that influence it, the next step is to clarify what practice actually entails. Because design thinking encompasses a number of methods and mindsets, it is exceedingly difficult to come up with a precise definition of its practice. *Is it a focus on applying specific methods? Is the practice of certain types of creative behaviors what should be captured?* Based on previous work much of the impact of design thinking is on dispositions (Royalty et al. 2012). Additionally, many organizations invest in design thinking because they want people to work in a certain way (Royalty et al. in press). This suggests that developing measures of practice should begin with how people approach solving problems in creative ways.

Seminal work in problem solving by Allen Newell and Herb Simon pointed to a model of how people approach complex challenges (Newell and Simon 1972). They broke problem solving into the following steps:

Intelligence—collecting enough information to identify the problem.
Design—developing multiple potential solutions.
Choice—evaluating the potential solutions and narrowing on one. Implementation—carrying out that decision.

Outcome—looking at the created solution and determining if it actually solves the problem.

Further work on creative problem solving (Adamson 1952; Csikszentmihalyi 1996) has led to more specific models that especially support approaching problems with very open-ended initial frames and/or outcome requirements. These models inspired a framework used to develop two different assessment tool pilots. The first captures how students approach general problems during the course of their d.school class (see the next section for details). This tool can also be used to describe the differences between various design thinking learning experiences that instructors create; this is a key variable identified above that impacts practice. The second tool assesses to what extent students can perform a short task that relies on them to use a disposition core to design thinking (i.e., prototyping). The following sections detail how these tools were piloted.

2.2 Assessment Materials

The assessment tools related to design thinking practices were piloted in two d. school courses. One was a course for advanced students—all of whom had previously taken at least one d.school course. The other was a course focused on using design thinking to drive organizational change. Both courses were a quarter long (10 weeks) and were relatively small with less than 15 students in each course. Our final population from the two courses was 18 students and four instructors.

Prior to the start of each course, students were given an online survey. The survey was comprised of the Creative Agency scale (Royalty et al. 2014) and six questions from the Innovation Self-Efficacy Scale (ISE) (Gilmartin et al. 2017). The entire ISE scale was not used because of survey length constraints and because many of the items were less relevant to these particular courses.

In addition to the scale items, a new assessment tool consisting of a short performance task focused on a particular disposition was given to the students through on online survey. The tasks are covered in more detail in Sect. 2. After the end of the course, students were sent a post-survey with the both the scale items and the same performance task.

Furthermore, in the final 2 weeks of each course, a researcher administered the Key Events Timeline activity that prompted students to delineate how they addressed salient challenges during the d.school course (see Fig. 1). Students were encouraged to include challenges they had encountered both inside and outside the class. The prompts did not ask students to focus on any specific part of the design process. Students were simply instructed to capture memorable events. Beneath each recorded event is a progression of how students approached, acted, and reflected on each listed challenge. This three-step model is a simplified version of problem solving models like Simon's.

After the conclusion of the course, each member of the teaching team filled out a similar timeline. The instructors were prompted to list the course events they

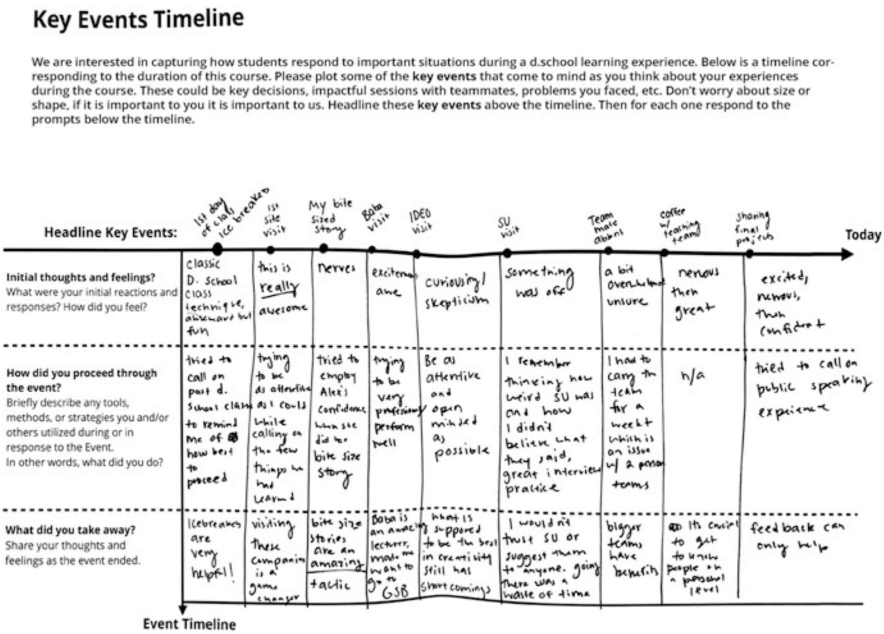


Fig. 1 Example of Key Events Timeline

believed were most salient to their students. For each event, instructors captured how they believed their students would respond to that moment or event.

2.3 *Data Analysis Approach*

The student and instructor timelines will be coded using a grounded theory approach (Corbin and Strauss 1990). Initial themes will be identified. Then a specific coding scheme will be created. The coding scheme will be useful for analyzing future versions of this tool. In addition, these student-generated codes will be compared to instructor-generated codes created during a previous study (Royalty et al. 2015). The interaction of these two coding schemes may provide new insight into the underlying teaching and learning model at the d.school.

The pre- and post-scale items will be analyzed to determine if there were changes during the course. Having a relatively small sample size (20 total students) may make that challenging though we anticipate this will provide a good foundation for larger-scale testing.

These measures will be iterated on in order to be used with students across multiple courses during the 2017–2018 academic year at the d.school and Columbia University.

3 Design Thinking Ecologies

3.1 Background

The final line of work continues the investigation of the role of design thinking within an organization. This work was born out of a need for organizations to compare and contrast how they teach and implement design thinking. Because each organization has its own innovation goals and reasons for using design thinking, it is difficult to translate learnings from one organization to another. For example, imagine a company begins to reduce the amount of design thinking trainings it runs in favor of supporting more project work using design thinking. Is this shift a "natural" progression of an organization, or is it dependent on other factors like innovation intent, industry, etc.? The framework of ecologies that we are developing highlights the context and environment in which design thinking is applied. This additional information allows organizations to share a much broader story of their evolution, which makes knowledge transfer more reliable.

The ecologies have been created through a design-based research model. An essential concept of this research model is iteration. Most of the iterations center around collecting data and testing previous models during regular gatherings of design thinking practitioners. Currently the ecology framework is on its third version. However, each version is based on the same theoretical underpinning.

Teresa Amabile's model for creativity in organizations (Amabile 1996) drives the ecologies. The model has two parts: Work Environment and Individual/Team Creativity. Both are necessary and are related. The environment has an impact on how creative teams can be. In turn, the creative work generated by teams drives innovation.

The first version was derived from Work Environment part of the framework (Fig. 2). The second version was built off the first. However, after testing the second version it became clear that participants recognized that the ecology had value beyond just comparing two or more companies. They saw it as an interesting tool to understand how different teams interact with the larger organizational environment. The third version now includes questions that elicit information about individuals and how they relate to broader systems that may or may not support design thinking work. Further iterations will continue to focus on describing the interplay between the environment and the individual or team.

3.2 Assessment Materials

Data for the ecologies were collected using a paper survey. The participants were primarily design thinking practitioners in industry, however, a few academic faculty also participated. The survey asked about: participation in design thinking trainings and design thinking projects, experience with design thinking, perceived value of

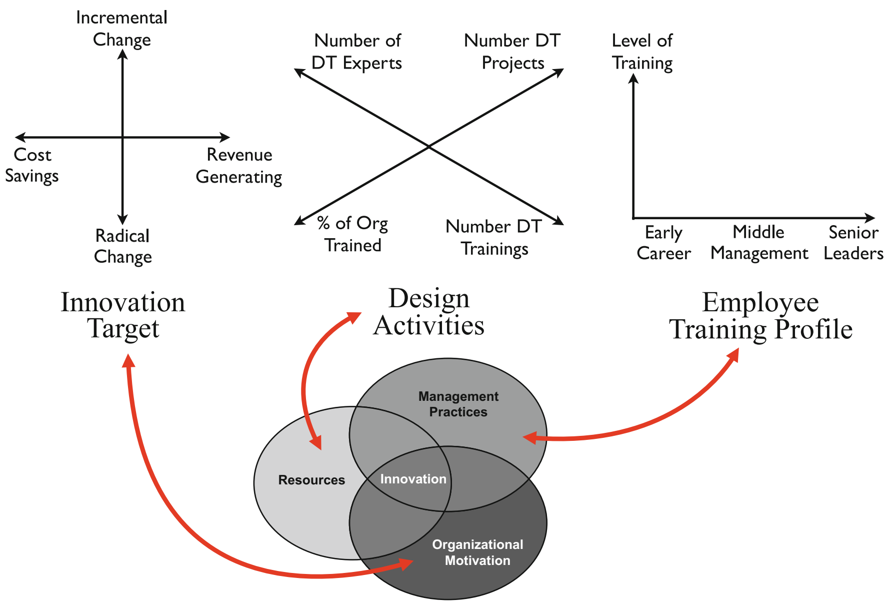


Fig. 2 Mapping Design Thinking Ecology 1.0 to Amabile's work environments

design thinking, and access to resources provided to support design thinking efforts. The surveys were administered during two different design thinking conferences.

In each case, participants were shown ecologies based on surveys collected at previous conferences. They were then asked to comment on the results of previous iterations of the ecologies. These comments heavily influence the continued iteration of this tool.

3.3 Results

The responses were sorted by company. Some companies currently only have one data point since they only had one conference participant. Most companies have two or more. In cases where a company has multiple data points, responses to each question were averaged. If the responses differed extremely, the respondents were contacted independently and asked to provide potential explanations for the discrepancy. Anonymity of all respondents was maintained.

Here are initial findings across five categories:

Organizational Diffusion of Design This captures where design thinking exists within an organization, specifically within business units. There were four general patterns that emerged (see Fig. 3):

- *Spread*—design thinking practiced in six or more areas.

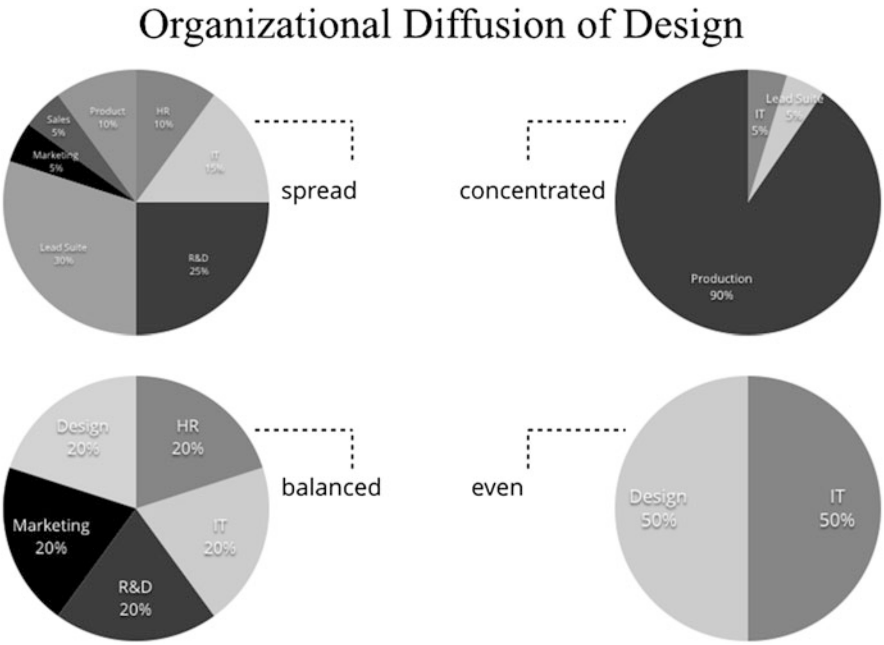


Fig. 3 Examples of organizations matching diffusion type

- *Concentrated*—one business unit responsible for 80% or more of the total design thinking.
- *Balanced*—design thinking fairly equally distributed between three to five business units.
- *Even*—design thinking split between two areas.

Number of Design Engagements This captures both the number of design thinking trainings and projects each organization performed in the past 6 months. Trainings were allowed to range from a 2-hour introduction to a multi day workshop. There is a break down of types of trainings but it is not presented here due to space constraints. Because most companies did not have an exact count of these engagement, the responses were organized/collapsed in to four categories: 0–10, 10–50, 50–100, and 100+. The vertical axis is the number of companies in our sample that fit into each category. An approximate total for trainings and projects is included (Fig. 4).

Design Training to Project Ratio This is the newest environmental measure. It measures the balance between trainings and projects. This theme has come up repeated in interviews with design thinking practitioners (Royalty and Roth 2016). In many cases, companies have intentionally reduced one of these two activities while increasing the other. This measure captures how many organizations fit into certain ranges.

Climate of Support for Design Mindsets This measure simply assesses how much both the overall organization and the respondents' specific team value design

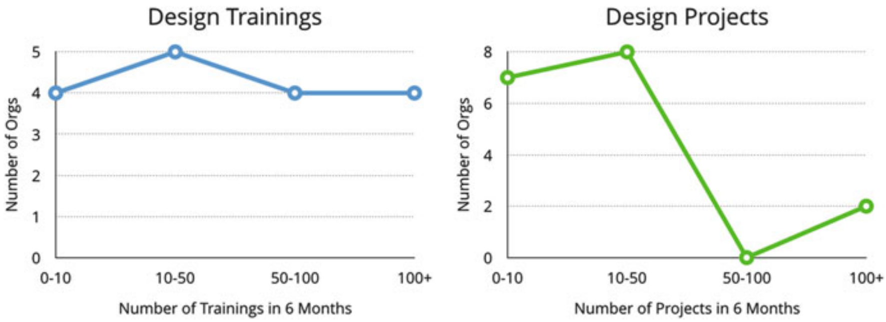


Fig. 4 Amount of trainings and projects conducted by companies

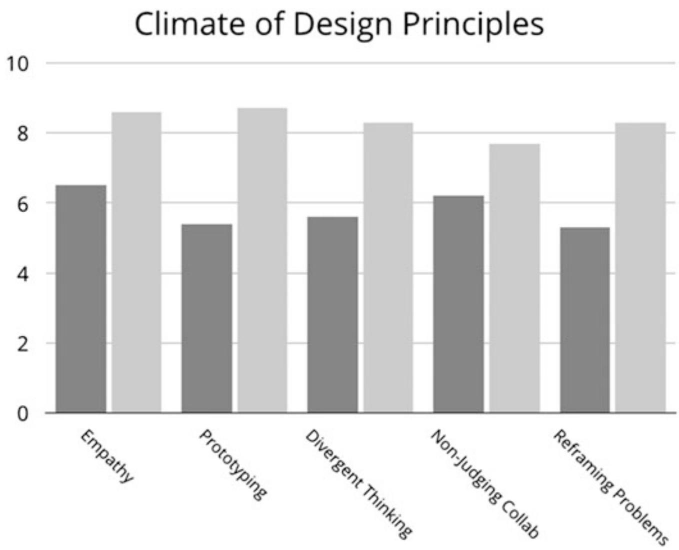


Fig. 5 Comparison of team and company values of design thinking

thinking. Respondents where asked how much their organization and their current team value: empathy, prototyping, divergent thinking, non-judgmental collaboration, and reframing problems. The chart below (Fig. 5) shows these data averaged across organizations. As one would expect, the team tends to value design thinking more then the organization.

There are still many more ways to analyze these data. But some initial results suggest a few things. One is that 17 companies combined to conduct over 3500 trainings and 650 projects within a 6-month timeframe. Even though two of the organizations accounted for roughly 50% of this activity, it still indicates the scale of which design thinking is being practiced in industry.

The ratio measure is particularly interesting, as many companies have actively worked to lower the number of trainings in favor of more projects. This often represents a shift in strategy from having an organization-wide culture of design thinking to using design thinking in a more targeted fashion to generate new products, services, or processes. A next step in our research will be to examine if more mature organizations have lower ratios. This analysis is not as straightforward as it sounds because design thinking might be mature in one business unit and nascent in another.

As we collect more data on specific business units and individual interactions with the larger environmental aspects, it will be possible to generate a more complete ecology for each organization. This will help lead to a process by which organizations can thoughtfully compare themselves to each other and engage in deeper and more useful sharing.

4 Conclusions and Future Work

This chapter suggests that there is a range of design thinking practice both in academic contexts and in industry contexts. Furthermore, the range is not necessarily along a single "weak to strong" strong axis. It is, in fact, much more complex and context dependent. Therefore if we want to understand ways to utilize design thinking it is important to develop measures that can detect multiple ranges of practice. Below are some examples of how different design thinking practices can range.

Design thinking ranges in how it is taught. Full courses, pop-up courses, and workshops are just a few of the multiple instructional formats used. The learning goals for these experiences can vary as well. There are several opportunities for further study. For example, looking more closely at how instructors design learning experiences should provide insight into precisely how these formats differ.

The student experience in a given course vaires. An initial inspection of the Key Events Timeline data shows that each student listed a unique set of critical moments in the course. Many timelines were completely different from one another. This suggests that instructors should provide a variety of experiences in order to reach as many students as possible. A next step would be to explore why certain students find particular events more critical than others.

Design thinking also ranges in how it spreads across an organization. The Organizational Diffusion measure shows that design can look extremely different from one company to another. It reminds us that labeling an organization as practicing design is a major oversimplification. *Does the whole organization practice it? Is it concentrated in one place?* As the ecology framework develops we will be able to draw better comparisons between organizations that use design in similar ways. Furthermore, there is an opportunity to be even more specific by mapping design practice within business units.

The final range we will highlight is how design thinking is used in an organization—specifically in terms of teaching and project work. Some organizations have a high training to project ratio. For others this ratio is low. This can indicate an extremely different purpose for design thinking. We suggest looking at how these ratios change over time. It may also be worthwhile to differentiate between different types of trainings and projects.

Our intent is that by identifying ranges of practice leaders looking to capitalize on design thinking will have a more nuanced perspective on how to implement and evaluate it. In other words, design thinking is not a single competency that you have or you don't have. It represents a range of different practices that can be improved and adapted if invested in thoughtfully.

References

Adamson, R. E. (1952). Functional fixedness as related to problem solving: A repetition of three experiments. *Journal of Experimental Psychology, 44*(4), 288–291.

Amabile, T. M. (1996). Creativity and innovation in organizations. *Harvard Business School, 5*(9), 396–239.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. Hoboken: Wiley.

Beckman, S., & Barry, M. (2007). Innovation as a learning process: Embedding design thinking. *California Management Review*, *Fall, 2002*, *50*(1).

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education, 24*(1), 61–100.

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons and evaluative criteria. *Zeitschrift für Soziologie, 19*(6), 418–427.

Csikszentmihalyi, M. (1996). *Flow and the psychology of discovery and invention*. New York: Harper Collins.

d.school. (2017, June 1). Retrieved from https://dschool.stanford.edu

Gilmartin, S. K., Chen, H. L., Schar, M. F., Jin, Q., Toye, G., Harris, A., Cao, E., Costache, E., Reithmann, M., & Sheppard, S. D. (2017). *Designing a longitudinal study of engineering students' innovation and engineering interests and plans: The Engineering Majors Survey Project. EMS 1.0 and 2.0 Technical Report*. Stanford, CA: Stanford University Designing Education Lab.

Hattie, J. (2003). *Teachers make a difference, What is the research evidence?*.

Hawthorne, G., Saggar, M., Quintin, E. M., Bott, N., Keinitz, E., Liu, N., et al. (2016). Designing a creativity assessment tool for the twenty-first century: Preliminary results and insights from developing a design-thinking based assessment of creative capacity. In *Design thinking research* (pp. 111–123). Berlin: Springer.

Jablokow, K. W., & Kirton, M. J. (2009). Problem solving, creativity, and the level-style distinction. *Perspectives on the nature of intellectual styles*, 137–168.

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of General Psychology, 13*(1), 1.

Kelley, D., & Kelley, T. (2013). *Creative confidence: Unleashing the creative potential within us all*. New York: Crown.

Kienitz, E., Quintin, E. M., Saggar, M., Bott, N. T., Royalty, A., Hong, D. W. C., et al. (2014). Targeted intervention to increase creative capacity and performance: A randomized controlled pilot study. *Thinking Skills and Creativity, 13*, 57–66.

Kolko, J. (2015). Design thinking comes of age. *Harvard Business Review, 93*(9), 66–71.

Köppen, E., Schmiedgen, J., Rhinow, H., & Meinel, C. (2016). Thisisdesignthinking.net: A storytelling-project. In *Design thinking research* (pp. 13–15). Cham: Springer.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, no. 9). Englewood Cliffs, NJ: Prentice-Hall.

Rauth, I., Köppen, E., Jobst, B., Meinel, C. (2010). *An educational model towards creative confidence.* 1st Proceeedings of ICDC, Kobe, Japan.

Richards, R. (1990). Everyday creativity, eminent creativity, and health: "Afterview"; for CRJ Issues on creativity and health. *Creativity Research Journal, 3*(4), 300–326.

Royalty, A., & Roth, B. (2016). Developing design thinking metrics as a driver of creative innovation. In *Design thinking research* (pp. 171–183). Heidelberg: Springer.

Royalty, A., & Sheppard, S. (2018). Mapping and measuring design thinking in organizational environments. In *Design thinking research* (pp. 301–312). Heidelberg: Springer.

Royalty, A., Oishi, L., & Roth, B. (2012). "I Use It Every Day": Pathways to adaptive innovation after graduate study in design thinking. In *Design thinking research* (pp. 95–105). Berlin: Springer.

Royalty, A., Oishi, L., & Roth, B. (2014). Acting with creative confidence: Developing a creative agency assessment tool. In *Design thinking research* (pp. 79–96). Heidelberg: Springer.

Royalty, A., Ladenheim, K., & Roth, B. (2015). Assessing the development of design thinking: From training to organizational application. In *Design thinking research* (pp. 73–86). Heidelberg: Springer.

Saggar, M., Quintin, E. M., Kienitz, E., Bott, N. T., Sun, Z., Hong, W. C., et al. (2015). Pictionary-based fMRI paradigm to study the neural correlates of spontaneous improvisation and figural creativity. *Scientific Reports, 5*, 10894.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. *The Nature of Creativity*, 43–75.

Making Use of Innovation Spaces: Towards a Framework of Strategizing Spatial Interventions

Marie Klooker, Martin Schwemmle, Claudia Nicolai, and Ulrich Weinberg

Abstract This chapter explores the use of space as a means of effectively fostering creativity and innovation in organizations more specifically, so-called spatial interventions. Spatial interventions refer to the strategic scope of actions that can be undertaken when and while using innovation spaces. We approach the topic of spatial interventions from different perspectives and shed light on crucial aspects of how innovation spaces can be used as a 'silent coach.' This approach takes into consideration the conceptual interplay of the strategic discourse, theoretical accounts of coaching practice and the process of using innovation spaces. We further introduce a framework for spatial interventions that helps to structure and analyze the use of space during a workshop. Finally, based on findings from a case study that was conducted in the newly created innovation space of a large company, we apply and expand this theoretical framework. The approaches and findings of this chapter support both strategists and practitioners and contribute to a deeper understanding of how to make use of space as a strategic tool.

1 Introduction

Designing effective workspaces for innovations in organizations is becoming the strategic focus of many organizational efforts. Several companies acknowledge the power that workplace design can exert on the way people think, behave, and interact. Thus, a lot of strategic efforts are undertaken to innovate space design. Even though the strategic intent is clear and the strategy is clearly articulated with regard to fostering and supporting efficiency, effectiveness, empowerment of innovation teamwork, development of new business designs as well as the expression of the organization's innovation values, research only provides very vague answers on how this strategy should be practiced in innovation spaces (Groves and Marlow 2016; Moultrie et al.

M. Klooker (✉) · M. Schwemmle · C. Nicolai · U. Weinberg
HPI School of Design Thinking, Hasso Plattner Institute for Digital Engineering, Potsdam, Germany
e-mail: marie.klooker@hpi.de; martin.schwemmle@hpi.de; claudia.nicolai@hpi.de

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_5

2007). We want to close this research gap by focusing on the actors' level and therefore the process of using innovation spaces. Innovation spaces are produced by its users interacting with the place and interacting among each other in the place. When focusing on the actors' level of innovation practitioners, we need to better understand what shapes the usage of micro-structures and structural elements within the innovation space, thus, the interactional patterns of place making. The question of "what to make use of" is connected with the question of "how to make use of."

This chapter outlines a framework for spatial interventions to be used by both strategists and practitioners in regard to understanding how the physical environment can be used in the wider context of organizational innovation, innovation capacity building, and creativity. Thus, the aim of this chapter is twofold. First, we want to provide the conceptual interplay between strategic intent and the use of innovation spaces. And second, we want to further delineate the strategic scope of actions that can be undertaken when using innovation spaces—so-called spatial interventions.

The structure of this chapter is as follows: in Sects. 2 and 2.2 we provide an understanding of our strategic approach and clarify our definition of strategy-as-practice for innovation spaces in use. We will focus on the interplay of people and interactions and clarify how tangible and intangible microstructures can be actively used as tools. In Sect. 3, we further introduce our conceptual framework of spatial interventions. This framework helps us to systematize the scope for actions and their effect on outcomes of innovation processes. Section 4 describes how these spatial interventions are used in a real setting by presenting results and findings from a case study. We close the chapter with a discussion in Sect. 5.

2 Innovation Spaces in Use: How Spaces Facilitate Organizational Innovation

2.1 The Concept of Strategy for Innovation Spaces: From Intended Strategy to Strategy-as-Practice

A lot of companies employ the strategy of innovation capacity building which is defined as designing appropriate organizational structures and processes, and thus implementing new business practices and workplace organization (Armbruster et al. 2008; Camisón and Villar-López 2014). Significantly changed organizational structures that will foster the implementation of advanced innovation policies, processes, and procedures based on deliberate strategic decision-making can be easily made visible by creating new tangible structures (OECD 2005). One such tangible structure can be an innovation space which companies implement as part of their innovation capacity-building activities. Accordingly, Moultrie et al. (2007) have pointed out that a clearly articulated innovation strategy in line with the intended design of dedicated innovation environments, physical workplaces and structures helps to understand *why* an organization wants to develop specific capabilities for enabling innovation. In

Fig. 1 Types of Strategies, based on Mintzberg and Waters (1985), own illustration

the following, we want to deepen the understanding of such innovation spaces as strategic tools and embed the use of space and spatial interventions within the strategic discourse.

Strategies in an organization form in different ways. Following Mintzberg and Waters (1985), a strategy can be considered as a continuum with two extremes—intended and realized strategies. While an intended strategy refers to leaders defining a strategy and thereby stating what the organization *should do*, the realized strategy focuses on what the organization actually *did*. Moultrie et al. (2007) build on this distinction by the notion of a "strategic intent" which is formed into a "realized intent" in the context of innovation spaces. The term they use, however, does not equal the more motivational idea behind strategic intent, as defined by Hamel and Prahalad (1989).

Building on the comparison of the intention and the realization of a strategy, two further types of strategies can be distinguished, deliberate and emergent strategies. Figure 1 summarizes the different types of strategy. A strategy is called deliberate if the strategy is realized as intended. Such a perfectly deliberate strategy requires three conditions to be met. The first two conditions refer to the intention of the strategy which, first, must be "articulated in a relatively concrete level of detail" and thereby leave no doubt about what is desired (Mintzberg and Waters 1985, p. 285). Second, because organizational innovation means collective action-taking, the intentions must be common for all actors (i.e., they must be "either shared as their own or else accepted from leaders") (ibid.). Third, the realization of the intended strategy must not be endangered by external forces, such as market or technology. In contrast to such a deliberate strategy, a strategy is called emergent if there is "consistency in action over time—in the absence of intention about it" (ibid.) The focus shifts from (intended) strategy, something an organization has, to something its organizational members do. In regard to the provision of a new workplace environment, its need to support different ways of thinking (divergent and convergent thinking) as well as different work modes of the innovation process (seeking inspirations, finding focus, creating ideas, developing prototypes and iterations) is crucial (Moultrie et al. 2007; Schwemmle et al. 2017). Consequently, even a delineated space strategy with the intent to enable and support innovation processes in a new manner needs to provide room for emergent elements.

The distinction between intended and emergent strategies is also reflected by the concept of strategy-as-practice, as introduced by Whittington (1996, 2003, 2006) which focuses on *how* a strategy emerges informally from managerial activities. In addition, Jarzabkowski et al. (2007, p. 7) distinguish between strategy and strategizing. They define strategy as "conceptualized as a situated, socially accomplished activity, while strategizing comprises those actions, interactions and negotiations of multiple actors and the situated practices that they draw upon in accomplishing that activity." Thus, the concept of strategizing helps us to shape the process of using innovation spaces designed to support and foster innovation development, teamwork, as well as creativity and design (Moultrie et al. 2007).

2.2 Strategy-as-Practice: Making Use of Microstructures in Innovation Spaces

Activities that are connected with the strategy of innovation capacity building within an organization encompass a variety of social activities. Therefore, we want to focus on the social activities that can be carried out in organizational innovation spaces (Whittington 2006). Setting up innovation spaces as a strategy creates a supporting environment to foster teamwork, innovation, creativity, and design on the micro-level (Moultrie et al. 2007). This level of structural organizational innovation that influences responsibilities, information flows, and cross-functional exchange on a macro-level has to be aligned with the level of procedural organizational innovation that effects activities, routines and practices (Armbruster et al. 2008). The strategy-as-practice approach, as introduced above, links the micro- and macro-perspectives on strategy as a social practice and enables us to better understand both micro-phenomena and the use of structural elements in their wider strategic and organizational context (Jarzabkowski et al. 2007).

Lipmanowicz and McCandeless (2014) have highlighted the power of making active use of microstructures and structural elements in shaping and guiding how groups interact and work together. They distinguish between tangible and intangible microstructures and structural elements (see Table 1). Tangible microstructures are defined as the physical spaces where innovation teamwork takes place and the tangible structural elements. These include tables, chairs, and resources that refer to all choices about how a space is arranged. This space can either contribute to or contradict the strategic intent. For example, the invitation to actively participate does not work well if none of the structural elements in the innovation space are allowed to be re-arranged by its users/participants.

However an approach to systematizing the repertoire of actions, tools and techniques, and of how to design the use of space in an intentional manner is still lacking. Also, the liberating structures that summarize potential actions that identify triggers to productive practices establish new patterns of behavioral habits suggest an "activity composing" approach based on the analysis of a specific innovation challenge (ibid., p. 86). In line with Simon Sinek's Golden circle (2011) and De Certeau (1984), we would like to shift the conceptual discussion from the why of innovation capacity

Table 1 Hierarchy and examples of microstructures and structural elements (Lipmanowicz and McCandeless 2014, p. 11)

	Tangible structures	Intangible structures
Micro-structures	Boardroom	Presentation/lecture
	Classroom	Managed discussion
	Meeting room	Status report
	Restaurant	Open discussion
	Office	Brainstorm
	Water cooler	Liberating structures
Structural elements	Large round table	Purpose/agenda
	Large rectangular table	Question
	Small table	Theme
	Chair	Seating arrangement
	Flip chart	Group configuration
	Post-its	Time allocation
	Projector	Standing instead of sitting
	Screen	Formal or informal

building as a strategy to the what and how of strategy-as-practice for innovation spaces in use. We will therefore address the following questions:

- What are the tools and techniques of strategizing the innovation spaces to be used?
- How can the tools and techniques of strategizing innovation spaces be used in practice?

In the following, we want to develop a framework of spatial interventions that will help innovation practitioners to make better informed decisions about what is, could or should be their repertoire of microstructures and the arrangement of structural elements when working in innovation spaces. We will also illustrate how tools and techniques of strategizing innovation spaces can be used and will share some findings about a case study that we conducted in a corporate environment where the company started to use a newly created innovation space.

3 Liberating Innovation Capacities: The Relational Concept of Place

3.1 Praxis, Practices, and Practitioners of Innovation Spaces

To systematize tools and techniques for the process of using innovation spaces, we will shed light on the interplay of people, spatial structures, and organizational context (Schwemmle et al. 2017; Klooker et al. 2016). The concept of strategizing to define the process of using innovation spaces sets the focus on people who—intentionally or unintentionally—*do* something and thus on the relational interplay between people and actions. Whittington (2006) proposes three elements as the nexus of strategizing: praxis, practices, and practitioners. Praxis is consequential for the strategic intent. It is considered as “all the various activities involved in the deliberate formulation and implementation of strategy” (ibid., p. 619). Practices refer to all shared routines and rituals of behavior, including procedures for thinking, acting and using things. Practitioners “are strategy’s actors, the strategists who both perform this activity and carry its practices” or, in other words, “the actors who shape the construction of practice through *who* they are, *how* they act, and *what* resources they draw upon” (Jarzabkowski et al. 2007, p. 11). Since this chapter aims at understanding the role of spatial interventions that are obviously initiated by people, we, in the following, will focus on practitioners as the actors of strategy.

We consider activities, workshops, and innovation project teams working in an innovation space as being a series of spatial interventions over time. This definition entails seeing the environment as an educator, as a third coach/teacher (cf. Strong-Wilson and Ellis 2007). Our definition of the spatial concept and what and how to make use of, it is a relational one. In this context, space is defined as the relation between objects and their environment, objects and objects, people and objects as well as between people and people. We define spatial interventions as the strategic scope of actions that can be undertaken when and while using innovation spaces and consider them as structural and procedural at the same time (Lefebvre 1974; Löw 2001). Structural interventions include changing the relationships between these elements (e.g., moving a table or taking a team outside). Procedural interventions include changes of work processes, procedures and routines that people display while interacting with each other in the environment (e.g., using the starfish method and lying on the ground instead of standing in front of a whiteboard while ideating). The differentiation between structural and procedural interventions is somewhat blurred because both intervention types represent two sides of one coin: Moving the set-up of furniture in the space (e.g., unifying different team spaces to one big space for sharing results or ideas) changes the structure and at the same time includes the procedural component of asking people to transform the space so that it is easier to share with the other team. Thus, the way we use space as a concept is not only simply a matter of distance; whatever we do to position ourselves in the space and as we move around will result in meaning-making and communication (Strong-Wilson and Ellis 2007; Lawson 2001) which might also lead to misinterpretation. The structural intervention of getting rid of furniture (tables and chairs) to let “creativity freely flow” can be interpreted as depriving the team of its resources. The procedural intervention of a coach who says, “now let go and think without any constraints about radical ideas,” without distancing himself from the team, can be interpreted as mistrusting the team to work on its own, even though this intervention was meant to give the team freedom to explore.

3.2 *Make Use of Innovation Spaces*

Finally, we want to address the question of how innovation practitioners can make use of the relational concept of space. Once an innovation space is set up, the place has been defined for an intended use by its planners, designers, and strategists. In line with this notion, the place becomes a locus of strategy. Strategy postulates a place that can be delimited as its own. Thus, the place postulates specific intentional use and thereby speaks to its users. Based on our previous study of a variety of innovation spaces during their process of creation and initial use, we were able to identify three core levels of structural interactions and speech-acts (Klooker et al. 2016). We differentiate between, first, spatial structures that provide resources such as materials, tools, education, time and space. Second, spatial structures that facilitate different work modes and interaction patterns according to the task(s) of the team (e.g. mobile high tables and whiteboard, human facilitators and process-orientation such as Design Thinking). Third, spatial structures that communicate the intention of the space such as the organization's motivation to foster teamwork and creativity (e.g. a rough and unfinished look and feel promotes doing, highlighting teamwork through displayed principles of team speak in the space).

These are also key qualities used to describe the qualities of team leadership coaches (Wageman et al. 2008; Hawkins 2013): A coach supports and helps the team to structure its way of thinking and working, and initiates a performative collaboration among the team members. Hackman and Wageman (2005, p. 269) define team coaching as "direct interaction with a team intended to help members make coordinated and task-appropriate use of their collective resources in accomplishing the team's work." To be more precise, team leadership coaching is about the development of team members capabilities to attain higher levels of accomplishments with regard to team innovation (DeRue et al. 2010; Rousseau et al. 2013). A team leadership coach sets, adapts, and changes the conditions of team work. Using the collective resources well, strengthening the team's functioning, and helping to shape the performance by leveraging existing knowledge and skills corresponds to the qualities of an innovation place being a communicator (motivational level), facilitator (consultative level), and provider (educational level) (Hackman 2002; Klooker et al. 2016). In essence, we therefore consider the space as a "silent" coach (neither a "silent" team member nor a "silent" leader), because the key characteristics of making use of microstructures in innovation spaces correspond to the qualities of team leadership coaching.

4 Spatial Interventions in Innovation Workshops

4.1 Roles of a Coach and Characteristics of a Situation: Theoretical Foundations

Following the notion of the space being a silent coach during a workshop, we shed light on two important and related theoretical foundations, First, we seek to understand the function a coach has to fulfil (over time) before and during a workshop. Second, we provide a better understanding of the workshop as a situation where coaching and thus spatial interventions take place. We then combine these two theoretical foundations—coach functions and characteristics of a situation—to develop our own framework for spatial interventions. As mentioned previously, we consider activities, workshops, and innovation project teams working in an innovation space as being a series of spatial interventions over time. A spatial intervention is defined as an activity that changes the structures and/or processes in a space and thereby de-routinizes the space (Sutherland 2013).

According to Hackman and Wageman (2005, p. 273), coaching functions "are those interventions that inhibit process losses and foster process gains." They propose three performance processes of a team and accordingly three functions of coaching which we—inspired by Simon Sinek (2011)—summarize as the Why, How, and What of coaching. These three dimensions of coaching also follow the competences of a group facilitator suggested by Stewart (2006) and are summarized in Table 2.

1. The *motivational* level of coaching minimizes free riding and builds shared commitment. It thereby supports the performance process of effort and addresses the question of what motivates the team, i.e., *why* it works together.
2. The *consultative* level of coaching aims at minimizing mindless adoption of performance routines and guiding next steps along the task requirements. Consultative coaching supports the performance strategies of a group and thus answers the question of *how* the team works.
3. The *educational* level of coaching is aimed at building the knowledge and skills of a team and thus balancing the contributions of team members. Therefore, educational coaching refers to the learned content, or: the *what*.

This threefold distinction of coaching roles corresponds to the three functional dimensions of an innovation space as communicator, facilitator and provider (Klooker et al. 2016). As a coach, a space therefore takes the following roles:

Table 2 Coaching roles (Hackman and Wageman 2005)

Coaching role		Aims at
Why	Motivational	Effort, commitment
How	Consultative	Performance strategies
What	Educational	Knowledge, skills

1. The space can motivate team members and the team dynamics (motivational communicator). For instance, changing positions or moving to other spaces for new work modes during the workshop can energize a team. Further, the spatial set-up (around a table, next to a whiteboard, . . .) can foster the team experience. In addition, the creation of a retreat space, which is a bit separated from other teams, allows a team to work without distractions and to focus, as well as to foster a sense of unity. In a similar way, a space can support the performance of single team members. A safe environment encourages them to think and act freely and creatively and thereby increase their motivation.
2. The space can further facilitate the workshop process (consultative facilitator). On a basic level, the space needs to provide all furniture and material necessary to run the workshop. On a more advanced level, the space can support different phases of a workshop. For instance, during a diverging phase where new ideas are generated, a light and inspiring surrounding might help, whereas a converging phase needs a more reduced place that enables focusing and concentration. If prototyping is included, the material which is provided can also inspire participants or lead them to using certain materials or a certain technology.
3. Lastly, the space can also take the role of an educator and knowledge provider (educational provider). First, it may provide information on the process or workshop agenda (e.g., through charts on the wall or elsewhere in the space). Second, it can be used to set a challenge-related workshop atmosphere. For instance, a space hosting a workshop in the field of vacation can have deck chairs, umbrellas, maps, and pictures to put the participants in the right mood.

We next elaborate on the situation and its characteristics. According to Belk (1975, p. 157), a situation "comprises a point in time and space" and works as a stimulus influencing a person (organism) and evoking a behavioral response. He differentiates the following five characteristics of a situation:

1. the physical surroundings of a person,
2. the social surroundings of a person, including other people and their characteristics and roles,
3. the temporal perspective, such as time since or to another event or situation,
4. the task definition underlying the situation,
5. and the antecedent states, such as momentary moods and momentary conditions a person brings into the situation.

Considering the space as a coach, these five characteristics of a situation can provide the following five insights to spatial intervention, which will also be discussed in more detail in the next section.

Physical Surroundings The space of a workshop should not only be considered as one big space, but rather as spaces within a space. That means, every team—and maybe even every team member—defines its own team space within the space and develops feelings of ownership during the workshop (Pierce and Jussila 2010).

Social Surroundings The spatial setup not only hosts a process and several teams, but also influences the team dynamics and thus the interaction within a team. For instance, the position of people at a table conveys their status within the group. Further, small spaces might force a more active interaction between team members.

Temporal Perspective In terms of a workshop, there are two levels of a temporal view on space. First, there is a before and a during the workshop (i.e. somebody can prepare the space in a specific way before the participants enter or the space can be changed together with them during a workshop). Second, it refers to the life cycle of the workshop and the team. For instance, team processes follow the phases of forming, storming, norming, performing, and each phase has other requirements that a space can reflect (Tuckman 1965).

Task Definition The task definition of a workshop refers to its purpose and the specific challenge. As explained above, the space can create an atmosphere that fits to the context of the challenge.

Antecedent States During a workshop, a coach has to consider the momentary moods of the team and can use the space and spatial interventions to react to them. For instance, a team lacking energy might be energized by going outside; a team having no new ideas might move to another space to be inspired.

Having introduced the theoretical foundations of coaching functions and characteristics of a situation, we now link both to create a framework of spatial interventions.

4.2 A Framework for Spatial Interventions

We organize our framework for spatial interventions along four dimensions: (1) Time, (2) Reference Object, (3) Initiator, and (4) Type and we now explain them in more detail.

The *Time Dimension* of a spatial intervention answers the question of when the interaction takes place. This can either happen before or during the workshop. An intervention done before a workshop creates an atmosphere for participants when they enter the room. This could either trigger a new mindset and work-mode through a set-up which looks totally different from the meeting rooms participants know. Or it could direct participants to the content of the challenge. For instance, participants of a workshop aiming at future developments could be framed as time travelers. They receive a ticket for a time machine as an invitation and enter a room with a Stars Wars-like atmosphere, including posters, space ships, fancy objects, and ambience music. An intervention during the workshop mainly differs from an intervention done before in (1) that participants can compare the before and after and are actively involved in either the decision and/or the execution of the invention and (2) that it does not have to be planned, but can happen spontaneously. For example, the team might decide to move all furniture aside and concentrate on protoype work on the floor. Or a coach reacts to the momentarily mood of the team and takes the team outside. It is important to further take into consideration the different phases during a workshop. In the following, we will introduce three different lenses for workshop phases that are relevant in the context of spatial interventions. First, innovation workshops mostly cover a diverging phase that involves the generation of new insights or new ideas and increases the information available. This is followed by a converging phase that reduces the information available (Kaner 2006). Since these phases require different skills and methods, the spatial requirements also differ. In a converging phase, for instance, a team needs a distraction-free space that supports the reduction and allows it to become focused. Second, during a workshop, a team encounters the phases of forming, storming, norming, and performing (Tuckman 1965; see above). This shift of focus from a more team-internal to a more output-oriented perspective can also be motivated by the space. Third, and linked with the second aspect, is the feeling of ownership for the space (Dawkins et al. 2017). Whereas, in the beginning, participants might feel as guests, during the course of a workshop, they more and more become psychological owners of the space (i.e., they consider the space as "theirs"). This increase in feelings of ownership can, for instance, lead to usage patterns reflecting a feeling of territoriality or of a stronger personalization of the space. It further moves the initiating focus from the coach to the team (also see below).

The *Reference Object Dimension* of spatial interventions answers the question of what changes during the intervention (i.e., what are the targets of an intervention—people or objects?). The obvious option is to change the spatial set-up or furniture within the space. However, the workshop participants and thus their interaction patterns with the space can also be considered as a spatial intervention. For instance, a participant who sits across the table facing a second team member displays a rather 'reciprocal confrontational' posture. By moving his position to the edge of the table, he changes his role to a 'consorting' position. This becomes even more of a 'collaborating' role if he decides to join the other team member on the same side of the table because he now "sees the world from the same perspective" (Lawson 2001, p. 135). If these two team members now decide to draw their attention closer to the whiteboard (i.e., their vertical working surface behind their table), this changes not only their perceived roles but also the patterns of interactions, with both team members now working collaboratively. The consideration of both space and user as possible reference objects of an intervention reflects our conceptualization of space as outlined at the beginning of this chapter. It also underscores the notion of a space simultaneously containing structural and procedural components.

The *Initiator Dimension* of spatial interventions focuses on who initiated the intervention. Especially at the beginning of a workshop, this might be the coach of the team. However, in an ideal scenario, the team realizes during the course of the workshop the potential of spatial interventions and, instead of the coach, team members internalize this knowledge and become initiators themselves. Thus, the trigger dimension also underlines the fact that the relevance and use of space during a workshop can, in addition to the process and outcomes, also be learnings of the workshop. (In this connection, we also refer to the thoughts on psychological ownership in *Time Dimension*).

Table 3 Dimensions of spatial interventions

Time	Reference object	Initiator	Type
When?	What changes?	Who initiates?	Which function?
• Before a workshop (initial setup) • During a workshop (different phases)	• Structures (objects) • Processes (people and their interaction with people/objects)	• Coach • Participant	• Motivation • Consultation • Education

Finally, the *Type Dimension* of intervention refers to its coaching role as outlined above. Here, we ask the questions: Does this intervention foster team motivation and performance (motivation), does it cater to the process and the workshop flow (consultation), or does it relate to skills and knowledge (education)? Examples for these functions have already been provided above.

The four dimensions of spatial interventions are summarized in Table 3. They not only help to reflect the role of space in workshops and show new possibilities, but also to actively plan spatial interventions. For instance, during a workshop, a coach might reflect on her activities and realize that she triggers most spatial interventions. A conclusion might be to push the responsibility more towards the team. Or, when preparing a workshop, a coach might combine the dimensions in order to create new interventions. A coach might then ask herself: What could be a good intervention for my team (1) during a workshop that (2) changes their way of interacting with the space, (3) initiated by me and (4) fostering the team's motivation?

By taking a theoretical view of coaching and a situation, and then by regarding four dimensions of spatial interventions, this section helps coaches to better understand and thus use the space at hand as a second coach before and during an innovation workshop. It further underlines the importance of the space for an innovation workshop.

5 Spatial Interventions in Practice

5.1 Innovation Space in Use: A Case Study Approach

To explore the strategic phenomenon of innovation capacity building by spatial interventions, we conducted a multi-case study with companies who were establishing innovation laboratories. The qualitative study entailed the entire process of innovation space design from strategic intent to realized intent of the innovation spaces in use (Klooker et al. 2015, 2016, 2017). Establishing an innovation laboratory within an organization can be considered a spatial intervention on the macro-level. In line with the previous sections we will focus on spatial interventions of actors and micro-structures within the established innovation space.

The formulation of an explicit strategic intent plays an important role for defining the initial design of the space needed (Klooker et al. 2015). The critical point in unleashing the innovation laboratory's full potential, however, turn out to be whether innovation techniques and practices for using such a space are either already known, and even internalized, or are supported and acquired by means of human facilitation (Klooker et al. 2017). It seems obvious that especially those users who are not familiar with such innovation spaces and/or processes (e.g. Design Thinking) need good facilitation. Yet, leadership and high-performance team research highlight that even experienced innovation teams work more efficiently with team leadership coaches by their side (DeRue et al. 2010; Rousseau et al. 2013). In the studied cases we identified various versions of the facilitation of effective use of space. Among these were facilitators who acted as hosts to introduce and explain how the space works, what it has to offer and how to use it, as well as connecting different project teams and individual users. In one case, a team formed to manage the lab. Fulfilling different roles, the team members acted as hosts, programmers of the space and mentors of the innovation teams using the space. We also found that more experienced team members or even entire innovation teams who were more familiar with the space and/or processes became facilitators, acting as role models for less experienced colleagues and by that triggering certain activities within the space. And of course, there are also assigned coaches explicitly facilitating innovation processes and conducting workshops. In this chapter, we will especially focus on the latter case—the interplay between workshop facilitators and the innovation space. To explore how assigned innovation coaches use space to support the work process of teams and foster their interaction with the space, we observed an entire workshop and analyzed the empirical data based on the following questions: What are the spatial interventions during the design thinking workshop? How are these strategized by the facilitators?

In the following, we will first introduce the methodological approach taken in the empirical study. Second, we will outline the general spatial set-up of the innovation space and the particular Design Thinking workshop conducted. Subsequently we provide exemplary findings that illustrate spatial interventions identified in practice. We will conclude this section by highlighting crucial aspects in regard to the strategy-as-practice of spatial interventions.

5.2 Methodological Approach

During a one-day introductory Design Thinking workshop conducted by two coaches for two teams in the creative space of a large company, we collected empirical data by means of non-participatory observation followed by short interviews with workshop participants. Two researchers were present during the workshop and documented their observations by means of field notes and photographs. The latter mainly served as illustration and enabled the researchers to later remember specific situations described within their protocols of field notes. Additionally, all participants completed a survey before and after the workshop. The aim was to gain insight into their perceived experience during the workshop and previous knowledge about Design Thinking and the innovation space. At the end of the workshop the participants' learnings and teams' experiences were reflected upon in a plenary session and documented by the two researchers present, in the form of individual notetaking. The empirical data collected during the workshop was categorized and analyzed according to the dimensions of the conceptual framework of spatial interventions.

5.3 The Innovation Space and Workshop Set-Up

The innovation space was newly created and set up for collaborative teamwork and creative sessions. The interior design was inspired by the HPI School of Design Thinking, Potsdam and provided space for two teams with approximately six team members who wanted to work in a conversational and/or collaborative mode. The space was equipped with two standing tables and approximately ten high chairs. The tables were in hexagon shape and had wheels to allow for flexibility. An open shelf contained boxes filled with basic prototyping material, post-its, and pens to support visual and interactive work. One entire wall was painted with white magnetic whiteboard paint to allow for visual teamwork, and a large screen provided the possibility to share digital material and presentations. Next to the workshop area was the kitchen, equipped with coffee machines and tea makers. The kitchen, though regularly used by the employees working on the floor, also provided a meeting space for informal conversations among innovation team members.

The agenda of the workshop was designed along the different phases of the Design Thinking process. Due to time constraints, the transitions between phases were not made explicit during the workshop, instead the overall frame was the division in the problem space (Understand, Observe, Point of View/Define) and the solution space (Ideation, Prototype, Test). The group of workshop participants totaled ten employees. The group was diverse in terms of the employees positions and levels of experience within the company, as well as their roles and responsibilities within departments and teams. Participation in the workshop was voluntary and based on interest in learning the Design Thinking method and getting exposed to its mind-set. All participants had little to no prior experience with Design Thinking, and most of them were using the innovation space for the first time during this workshop, or had so far not used it for collaborative teamwork and innovation projects but only as a regular meeting space. Two experienced Design Thinking coaches led the workshop. Each one accompanied a team throughout the entire workshop.

5.4 *Spatial Interventions in Practice: Applying the Framework*

Exploring the Type Dimension of Spatial Interventions Overall, we identified various spatial interventions of all three *types* (motivational, educational, consultative) throughout the workshop. The two coaches took turns providing content and methodology related inputs (educational interventions) in short plenary sessions with both teams at the beginning of the workshop and in each phase of the design thinking process. This forced the synchronization of both teams and provided opportunities to share questions and doubts. Throughout the process and within each team the coaches' role was mainly to facilitate how to work (consultative interventions). Teambuilding activities in the beginning of the workshop, informal conversations during the break and sharing outcomes enforced the experience's value for all participants (motivational intervention).

Exploring the Time Dimension of Spatial Interventions Before the workshop started the coaches adjusted the set-up of the innovation space according to specific needs regarding content and agenda. Due to the fact that the space was intentionally designed as creative space with Design Thinking furniture, only little transformation was needed in preparation for the initial set-up. Two team spaces were created by moving the flexible tables. Both team spaces consisted of a high table in hexagon shape and five high chairs. One team was placed close to a wall that served as magnetic whiteboard. The other team space was equipped with one mobile whiteboard. The wheels on the high tables were fixed so that they would not move when leaning against them. The high chairs were stacked in the center between both team spaces with the intention to force the participants to stand at the beginning. The front of the space was left free for plenary presentations and inputs. On both tables, basic materials (post-its and pens) were spread out to invite participants to use them freely. At the beginning of the workshop the group was divided into two teams. The teams were formed according to the tables the people were standing at, disregarding existing hierarchies between participants.

Throughout the workshop the coaches have used different interventions to introduce the flexibility of the space. Consequently, the teams used the potential flexibility of their spaces differently. Before Team 1 started with the first task, the coach explicitly highlighted the flexible nature of the space by telling the team to use and move everything as needed (consultative intervention). Through this he empowered the team members to take action themselves and independently of him. The teams immediately became active in adjusting their space to their needs. Team 2 immediately started with the first task without waiting for further instructions by the coach. The coach did not intervene by explaining how to use the space and left it to the team to discover the furniture's flexibility. Our observations showed that Team 2 did not move the furniture throughout the entire workshop. Reasons may be that the flexibility was not as intuitively discovered due to the fact that the wheels on the table were all fixed at the beginning even though they had seen the other team moving its furniture. The coaches did not change the set-up during the workshop. Both coaches were experienced in conducting workshops in spaces with similar style and interior. Yet, they interacted little with the furniture and did not perform any major adjustments or rearrangements when preparing the initial set-up. Therefore, they remained neutral regarding the specifics of the space. This might add a reason why they did not trigger changes in the spatial set-up during the workshop.

Exploring the Initiator Dimension of Spatial Interventions We observed Team 2 as more active during the needs-finding-phase (problem-phase) and becoming less active during the second part of the workshop (solution-phase). While the team remained seated throughout the entire ideation session, the coach constantly moved between the whiteboard wall and the team to place their post-it's with ideas. The whiteboard wall was not movable; however, the table could have been moved closer to the wall or the team encouraged to stand up and re-group closer together in front of the whiteboard wall. At the same time, Team 1 did not change the spatial setting of the table and the chairs but changed its own spatial setting, moving themselves between table and whiteboard in order to work closer together. Additionally, the coach initiated and supported this intervention by moving into the background to leave the participants—physically—the space to work more independently. As a result, this spatial intervention, as initiated by the team itself and supported by the coach during the workshop, was successful in motivating the team to create more ideas.

Exploring the Reference Object Dimension of Spatial Interventions Team constellations changed throughout the workshop for Team 2. All 6 team members were present for the needs finding-phase, and all team members spread equally around the hexagon table with relatively large distances between individuals. One participant had to leave the workshop during the ideation phase. We observed how the thus created empty space at the team table affected the team dynamics: The remaining team members had to switch their position to avoid an otherwise awkward division of the team. These movements activated especially one team member who changed from a rather passive mode into one of taking initiative. He stood up and started moving between the whiteboard wall and the table, putting up post-it's with ideas from himself and his team mates. The increased dynamics seemed to motivate other team members to get up from their chairs and become more active as well. We observed how the increased activity also led to an increased amount of ideas created by the team. At the same time the coach moved more into the background and even left the team alone for a short period to provide space for the members to work on their own. Hence, the coach noticed that the team was finally able to interact with the space without her support.

Exploring Interrelations of Multiple Dimensions of Spatial Interventions: Time, Initiator and Reference Object Taking into consideration the entire workshop we found that the experiences of Team 1 and Team 2 differed in regard to different levels of activity throughout the process. Team 1 started off in a rather active mode, taking initiative to interact and make use of the space. The team members continued with a more or less same level of activity until the end of the workshop. Team 2 on the other hand, was rather passive in the beginning and the process seemed much tougher. This only changed in the second half of their ideation, when the team dynamics changed and the level of activity and ability of using the space shifted strongly (see above). After the ideation phase the team moved to the prototyping material that was placed a few meters away and available for both teams. The team was immediately inspired to prototype its idea. It seems as if the experience of being able to shift from a passive mode into an active one, and at the same time from feeling unproductive and uncreative into producing many good ideas, strengthened the members creative confidence. And this again led the spatial set-up, in terms of prototyping material, facilitate their process. All team members were motivated as they expressed their excitement about the selected idea and seemed confident about their overall outcome. Team 1 who had been actively interacting with the space throughout the entire workshop, seemed rather hesitant about building a prototype. After the ideation session, they also moved to the prototyping material but then decided that they first need to discuss again what and how to build before starting. Compared to Team 2 they seemed less motivated and excited about prototyping. And the mindset shift from creating many ideas to selecting only one to be built in the end, seemed more difficult for them. Both teams presented their prototypes, however Team 2 seemed less attached to their prototype and more open to possible iterations, based on feedback from the other participants. The iterative process, and its consequence of being able to define an outcome further based on feedback from the outside, is a crucial factor of Design Thinking projects.

This example shows that despite the differentiation of dimensions in regard to spatial interventions, in practice all dimensions are interdependent and the emergence of spatial interventions throughout the workshop cannot be planned.

5.5 Emergent Strategies for Spatial Interventions

The examples provided above highlight the differences of spatial interventions in regard to the two teams observed during the workshop. Taking into consideration the entire workshop, we may conclude that teams and their dynamics differ and therefore need different coaching strategies. Team 1 was generally more active and one team member already knew the space well. The coach's focus was on how the team should work: interactively and flexibly. His intention was mainly consultative, i.e. he aimed at enabling the team to apply Design Thinking as a method. The coach empowered the team already at the beginning of the workshop, encouraging them to interact with the space and to reposition the interior according to their own needs. The other team however, already started with the first activity without waiting for the coach's instructions. Team 2's coach allowed team members to stay in the flow experience and when they became rather passive she did not consult them on how to work differently but supported their needs, for example by moving between table and wall to put up their post-its, instead of forcing members to get up. Her focus at that point in the process was on motivating the team through a positive experience. Hence, an active and experience-oriented coaching strategy was applied. However, the coach's strategy also allowed for a crucial and spatial intervention to emerge from within the team, resulting in a strong shift in team dynamics. The "empty space" created by the missing members triggered remaining team members to become more active themselves, allowing the coach to leave the team to continue working indepndently. This in turn enabled the team to experience its own ability to create valuable outcome without the coach's close support. If the coach would have followed her initial coaching strategy—staying closely with the team and supporting it by taking over the participants' tasks—she might have inhibited their performance.

Building on the conceptual framework of spatial interventions introduced in the previous section of this chapter, the findings of our empirical study go beyond an illustration of the identified four dimensions and lead to the following conclusions. Taking into consideration the very different team experiences during the same workshop set-up and within the same initial space shows that different coaching strategies and team experiences are legitimate. Which one to choose and to what extent the actual spatial set-up becomes the 'co-coach' during a Design Thinking workshop depends on various factors that call for different strategic approaches. First, the gap between prerequisites of the general spatial set-up and degree of transformation needed before and during the workshop "sets the tone" and defines the starting point of any intervention (having to transform a hotel venue into a creative space vs. a well-equipped and intentionally designed creative space, and having loads of resources and large space available to set-up everything in the beginning vs. having to improvise throughout the workshop and rearrange the space according to different process steps). Second, the overall intention of the workshop experience serves as guide for the coach in regard to defining a strategy for spatial interventions (e.g. learning how to work in the Design Thinking mode vs. having a positive experience that motivates participants to learn Design Thinking vs. outstanding outcome). And third, the actual spatial interventions to be applied depend to a large degree on the people, on the level of individual team members (e.g., familiarity with the space, active and passive personalities, motivational nature etc.) and on the level of the team as such (e.g., team dynamics and mood, hierarchy etc.) and their interrelations (e.g., team dynamics). Whereas the first and second factors can and should be strategically planned in advance, the latter is more complex and often an often unknown before the start of a workshop. Independent of the initial strategic intention of a workshop, and therefrom resulting strategies for using the power of space as a second coach, an emergent strategy from the coach's perspective is needed in order to apply effective spatial interventions throughout a Design Thinking workshop. This subsequently allows the interplay of participants and spatial structures during the process of using a space.

6 Connecting the Dots: Strategizing Spatial Interventions

Focusing on innovation spaces in use, the previous sections help to better understand the power of innovation spaces as strategic tools to foster innovation and creativity. By means of exploring the domain of spatial interventions we contribute to intentionally designing the process of using innovation spaces. The different sections within this chapter examine (1) the integration of the topic of spatial interventions within the strategic discourse with the focus being on practitioners as the actors of strategy, (2) systematizing spatial interventions by conceptualizing a framework for them and thereby allowing an exploration of the role of space as the silent coach, and (3) looking at spatial interventions in practice to identify and explore their nature beyond the theoretical discourse. Each of the previous sections provides a different perspective on spatial interventions and sheds light on various relevant aspects. In this section, we aim to tie these areas together by connecting the dots towards a framework of strategizing spatial interventions.

We started our exploration by looking into the conceptual interplay of strategy and space and shifted the discussion from intended to emergent strategy, hence strategy-as-practice (Whittington 1996, 2003, 2006). This concept links the micro- and macro-perspectives on strategy as a social practice enabling us to better understand both micro-phenomena and the use of structural elements in their wider strategic and organizational context (Jarzabkowski et al. 2007). In line with our applied understanding of space as a relational concept (Löw 2001; Lefebvre 1974) we then introduced the concept of strategizing (Jarzabkowski et al. 2007). Strategizing refers to actions of multiple actors and their practices applied in accomplishing activities. The relational interplay of people and actions within an innovation space consists of intentional or unintentional *doing* that we refer to as spatial interventions. We understand spatial interventions here as activities that change the structures (distances, relationships between elements) and/or processes (work process, procedures, routines) in a space and thereby de-routinize the space (Sutherland 2013).

Linking the topic of spatial interventions to the practice of coaching innovation teams helped us to understand the role of space (and the people acting within it) as a 'silent' coach. Based on our understanding of a workshop as a situation where coaching and thus spatial interventions take place, we have looked into the function a coach has to fulfil before and during a workshop (and over time), as well as the threefold distinction of coaching roles. Building on this we created a framework of spatial interventions organized along the following four dimensions: (1) Time (when does it take place?), (2) Reference Object (what/who is moved?), (3) Initiator (who/what intervenes?), and (4) Type (what kind of intervention?). The latter also follows the three previously identified functional dimensions of an innovation space acting as communicator, facilitator and provider (Klooker et al. 2016).

The systemization of the strategic scope of actions within innovation spaces and during a workshop (1) helps innovation coaches to reflect on the role of space and how they do or could make use of it, it (2) may serve as tool to plan and formulate a coaching strategy in regard to spatial interventions, and (3) underlines the role of space for innovation workshops in general.

As part of a larger case study, we applied the conceptual framework as a tool to analyze spatial interventions observed during an introductory Design Thinking workshop conducted in the newly created innovation space at a large company. Besides illustrating the usefulness of the framework as a reflection tool for our observations, we generated insights that go beyond the conceptual frame. The analysis of our empirical data about spatial interventions in practice showed that despite the differentiation of dimensions, the spatial interventions are interdependent. Consequently, their oftentimes unintentional occurrence depends on various factors. As such their application cannot be planned deliberately, based only on the conceptual framework. Our observations during the workshop show differences in regard to team experiences and coaching strategies applied, yet, a similar outcome reached. This led to the conclusion that there is a need for emergent strategies of innovation coaches in regard to spatial interventions. To what extent the actual spatial set-up can and will become a 'co-coach' during a Design Thinking workshop depends on various factors that call for different strategic approaches. Further research with a larger empirical sample may focus on creating a catalogue of strategic approaches in coaching Design Thinking. However, within the limits of our data, already preliminary conclusions in regard to necessary spatial interventions could be drawn. First the initial set-up within a workshop and, related to this, the degree of necessary changes for a specific workshop are crucial factors. Second, a defined strategic intention of a workshop guides emerging strategies of coaches throughout the workshop. Third, the diverse nature in regard to team dynamics and individuals and the resulting specificity of every team is an important criteria for spatial interventions. In our case this turned out to be the unknown factor and/or effective change agent. As highlighted previously, it is crucial to note that only the first and second aspects can (and should) be strategically planned in advance, while the complexity of the latter calls for emergent strategies as opposed to deliberate ones.

Hence when looking at innovation spaces in use and the spatial interventions occurring throughout the process of using such a space, we need to apply the concept of strategizing to make sense of them at all. In practical terms this means that strategies to unleash the potential power of space in fostering innovation and creativity during the process of using the space (by means of spatial interventions initiated by innovation coaches and teams or by designed tangible structures) cannot necessarily be realized as planned if formulated and practiced deliberately. Facilitators therefore need to be able to strategize spatial interventions—in other words, to translate the strategic intent of a space or a workshop into emergent strategies. Having linked the conceptual framework for spatial interventions provided above to the strategic discourse and applying it to practice, this chapter contributes to the practice of unleashing the power of space in regard to innovation capacity building by again shifting the focus of spatial interventions from the '*what*' to the '*how*' and thereby providing the basis for a framework of strategizing spatial interventions.

References

Armbruster, H., Bikfalvi, A., Kinkel, S., & Lay, G. (2008). Organizational innovation: The challenge of measuring non-technical innovation in large-scale surveys. *Technovation, 28* (10), 644–657.

Belk, R. W. (1975). Situational variables and consumer behavior. *Journal of Consumer research, 2* (3), 157–164.

Camisón, C., & Villar-López, A. (2014). Organizational innovation as an enabler of technological innovation capabilities and firm performance. *Journal of Business Research, 67*(1), 2891–2902.

Dawkins, S., Tian, A. W., Newman, A., & Martin, A. (2017). Psychological ownership: A review and research agenda. *Journal of Organizational Behavior, 38*, 163–183.

de Certeau, M. (1984). *The practice of everyday life*. Berkeley: University of California Press.

DeRue, D. S., Barnes, C. M., & Morgeson, F. P. (2010). Understanding the motivational contingencies of team leadership. *Small Group Research, 41*(5), 621–651.

Groves, K., & Marlow, O. (2016). *Spaces for innovation: The design and science of inspiring environments, Fram3*.

Hackman, J. R. (2002). *Leading teams: Setting the stage for great perfomances*. Boston: Harvard Business School Press.

Hackman, J. R., & Wageman, R. (2005). A theory coaching of team. *The Academy of Management Review, 30*(2), 269–287.

Hamel, G., & Prahalad, C. K. (1989). Strategic intent. *Harvard Business Review, 67*(3), 63–78.

Hawkins, P. (2013). *Leadership team coaching: Developing collective transformational leadership* (2nd ed.). Kogan: Page.

Jarzabkowski, P., Balogun, J., & Seidl, D. (2007). Strategizing: The challenges of a practice perspective. *Human Relations, 60*(1), 5–27.

Kaner, S. (2006). *Facilitator's guide to particpatory decision-making* (2nd ed.). San Francisco: Jossey-Bass.

Klooker, M., Matzdorf, S., Nicolai, C., Boettcher, L., & Trost, A. (2015, December). The importance of strategic intent in developing innovation space. In *ISPIM innovation symposium (p. 1). The International Society for Professional Innovation Management (ISPIM).*

Klooker, M., Nicolai, C., Matzdorf, S., Trost, A., von Schmieden, K., Böttcher, L., & Weinberg, U. (2016). On creating workspaces for a team of teams: Learnings from a case study. In *Design thinking research* (pp. 67–84). Cham: Springer.

Klooker, M., Nicolai, C., & Uli, W. (2017, June). The innovation laboratory as tool for unleashing innovation capacity in a large organization. In *24th Innovation and Product Development Management Conference (IPDMC).*

Lawson, B. (2001). *The language of space*. London: Taylor & Francis.

Lefebvre, H. (1974). *The production of space*. Oxford: Wiley-Blackwell.

Lipmanowicz, H., & McCandeless, K. (2014). *The surprising power of liberating structures: Simple rules to unleash a culture of innovation*. Liberating Structure Press.

Löw, M. (2001). *Raumsoziologie, suhrkamp*.

Mintzberg, H., & Waters, J. A. (1985). Of strategies, deliberate and emergent. *Strategic Management Journal, 6*(3), 257–272.

Moultrie, J., Nilsson, M., Dissel, M., Haner, U.-E., Janssen, S., & Van der Lugt, R. (2007). Innovation spaces: Towards a framework for understanding the role of the physical environment in innovation. *Creativity and Innovation Management, 16*(1), 53–65.

OECD. (2005). The measurement of scientific and technological activities – Proposed guidelines for collecting and interpreting technological innovation data. European Commission/Eurostat.

Pierce, J. L., & Jussila, I. (2010). Collective psychological ownership within the work and organizational context: Construct introduction and elaboration. *Journal of Organizational Behavior, 31*, 810–834.

Rousseau, V., Aubé, C., & Tremblay, S. (2013). Team coaching and innovation in work teams. *Leadership & Organization Development Journal, 34*(4), 344–364.

Schwemmle, M., Nicolai, C., Klooker, M., & Weinberg, U. (2017). From place to space: How to conceptualize places for design thinking. In *Design thinking research* (pp. 275–298). Cham: Springer.

Sinek, S. (2011). *Start with why: How great leaders inspire everyone to take action, Portfolio*.

Stewart, J.-A. (2006). High-performing (and threshold) competencies for group facilitators. *Journal of Change Management, 6*(4), 417–439.

Strong-Wilson, T., & Ellis, J. (2007). Children and place: Reggio Emilia's environment as third teacher. *Theory Into Practice, 46*(1), 40–47.

Sutherland, I. (2013). Arts-based methods in leadership development: Affording aesthetic workspaces, reflexivity and memories with momentum. *Management Learning, 44*(1), 25–43.

Tuckman, B. W. (1965). Developmental sequences in small groups. *Psychological Bulletin, 63*, 384–399.

Wageman, R., Nunes, D. A., Burruss, J. A., & Richard Hackman, J. (2008). *Senior leadership teams: What it takes to make them great*. Boston: Harvard Business School Press.

Whittington, R. (1996). Strategy as practice. *Long Range Planning, 29*, 731–735.

Whittington, R. (2003). The work of strategizing and organizing: For a practice perspective. *Strategic Organization, 1*(1), 117–125.

Whittington, R. (2006). Completing the practice turn in strategy research. *Organization Studies, 27* (5), 613–634.

Part II
Exploring the Digital Potential: Teaching, Research and Organizational Approaches

An Iterative Approach to Online Course Design: Improving a Design Research MOOC

Karen von Schmieden, Lena Mayer, Mana Taheri, and Christoph Meinel

Abstract How can design thinking be taught in a Massive Open Online Course (MOOC)? The research team took an iterative approach at designing an online course about the design research phase and built a prototype MOOC for testing. We applied three measurement tools (Course Evaluation Survey, Skill Confidence Rating, Qualitative Interviews) and categorized the collected feedback. This process resulted in 57 iteration tasks, which were implemented in the public version of the MOOC. From over 5000 participants, 3034 learners participated actively in the public MOOC. 84.75% of survey participants were satisfied with the MOOC. In this chapter, we describe the iteration process and first results from the public MOOC.

1 Introduction

With the increasing use of digital resources and the integrating of the internet into pedagogical formats, more and more universities as well as private providers turn to internal online courses or Massive Open Online Courses (MOOCs). The number of MOOC offers is rising. As MOOCs have progressed, questions about the effectiveness and user-friendliness of the online courses emerge. In this chapter, we lay out how we applied a user-centered, iterative approach for refining a MOOC on design research. We first built a MOOC prototype (protoMOOC) and then tested it with a closed group of learners. For the protoMOOC testing, we applied a mixed-method approach. We gathered user feedback and evaluated it with three different measurement tools. From there, we categorized the feedback and formulated actionable tasks. By executing these tasks, we improved the online course and ran it publicly from August 2017 over a period of 5 weeks.

K. von Schmieden (✉) · L. Mayer · M. Taheri · C. Meinel
Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany
e-mail: Karen.Schmieden@hpi.de; Lena.Mayer@hpi.de; Mana.Taheri@hpi.de; meinel@hpi.de

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_6

2 ProtoMOOC Set-Up

Design thinking is a human-centered approach often used to identify and solve complex problems in society. MOOCs provide an opportunity to fulfill the growing demands for design thinking (online) formats. This helps in teaching design thinking on a large scale.

To collect user feedback, we followed an explorative approach by designing a protoMOOC. We ran and tested the protoMOOC for three reasons. First, we planned to investigate how the new online format for teaching and learning design thinking would be received by novice learners in comparison to experienced design thinking practitioners. Second, we wanted to test the content in a closed environment to iterate and improve the format, design, and applicability of the course. Third, we wished to test the course with a manageable size to understand how to handle online learners (specifically, their requests and complaints) before scaling the course to a larger public MOOC with international learners. In this chapter, we focus on the feedback-based advancement of the protoMOOC and describe the process of iteration.

We designed the protoMOOC based on the several research findings:

- We assessed best practice examples of existing design thinking MOOCs (Taheri and Meinel 2015).
- We followed recommendations by Siemens and Tittenberger (2009) and applied the *Seven Principles for Good Practice in Underground Education* (Chickering and Gamson 1987) to our instructional design.
- We evaluated and summarized learnings from testing modules of the educational material with a student group.
- We defined clear and measurable learning outcomes, based on the classification of learning outcomes by Kraiger et al. (1993).

One of our main decisions was to create a MOOC that targets skill-based learning and an individual learning experience. This enables learners to acquire skills of the design thinking process independent of the work and speed of other learners. In this way, we wanted to create a design research MOOC that is user-friendly, easy to approach, and has a low entry threshold.

Our second main decision was to break the design thinking process down into three separate skill-based MOOCs. For this we turned to IDEO.org's categorization of the human-centered design process. IDEO divides the process into three phases: Inspiration, Ideation and Implementation. Inspired by this, we chose to divide the process into the following phases: *Finding Inspiration*, *Creating Solutions*, and *Validating with Users*.

Our protoMOOC *Inspirations for Design* represented the first phase of "finding inspirations," in which we addressed the two basic skills of *Observation* (being attentive to one's surrounding) and *Qualitative Interviewing* (planning and conducting an interview, and identifying surprising findings to interpret). The learning objectives were defined according to Bloom's taxonomy (1994).

To convey the MOOC content, we utilized a variety of learning modes, e.g. videos, exercises, or peer assessment as implemented on openHPI (Staubitz et al. 2016). We provided a detailed description of the protoMOOC in the previous volume of *Design Thinking Research* (Taheri et al. 2017).

3 ProtoMOOC Measurement Tools

To gather feedback on the protoMOOC, we created various touch points with users, including: pre and post course evaluation surveys (CES), skill-confidence rating (SCR) and 16 qualitative interviews with interested learners from various backgrounds.

3.1 Course Evaluation Survey (CES)

The purpose of the CES was to compare learners' course expectations and satisfaction prior to course start and after course completion, as well as gathering data on basic demographics. The two surveys bracketed the course (see Fig. 1).

3.2 Skill Confidence Rating (SCR)

We introduced the skill confidence rating for the protoMOOC as a survey to assess the impact of the learning content on the learners' perception of their own skill development (introduced in the previous volume of *DTR*). The SCR consisted of two pre and two post ratings. Together, pre and post SCR bracketed the *Observation* (week 2) module and the *Qualitative Interviewing* (week 3) module. Learners had to indicate their confidence with skills central to these methods before (e.g., *How confident do you feel about conducting qualitative interviews*) and after taking the module (e.g., *How confident do you feel about conducting qualitative interviews after taking this week's learning module on Qualitative Interviewing*). The post ratings also included a question about what learners missed during the preceding module. They could write their answer in an empty text box. This question item helped to gather prompt feedback on the learning content. Besides, changes in skill confidence could be tracked back to the didactic design and unfulfilled learner needs.

The interpretation of the SCR allowed the course instructors to

(a) understand the initial level of participants' perception of their skills
(b) evaluate the effectiveness of course content (e.g., analyze the changes in skill confidence within a learning module).

3.3 Qualitative Interviews

Apart from gathering quantitative feedback in different surveys, we conducted in-depth qualitative interviews after the course that led to valuable feedback to iterate. We conducted 16 interviews with proto MOOC learners who agreed to being interviewed in the post CES. Interviewees were categorized as design thinking novices, design thinking experts, and international learners. All interviewees were asked about their personal experience with the course content, learning modes, platform features (e.g. peer review system, forums), participant activity and their device choice. The question set for design thinking experts included additional questions about feedback on design thinking methodology and didactic design. Finally, interviews with international learners were targeted at evaluating how the course content and design resonated with international learners in terms of cultural inclusivity.

Figure 1 illustrates at which times the evaluation tools were applied.

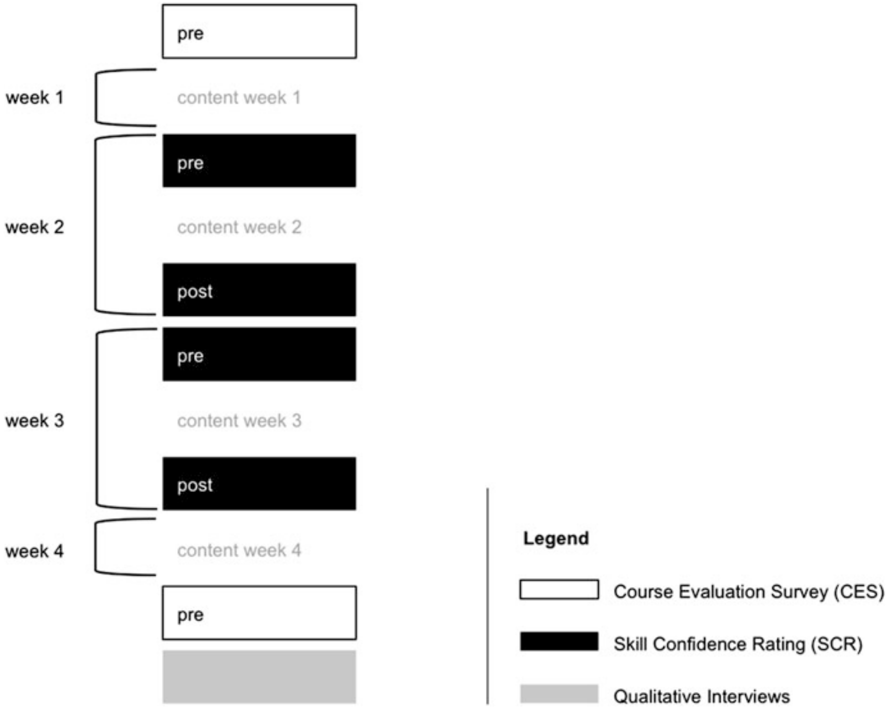


Fig. 1 Position of measurement tools within the protoMOOC

4 ProtoMOOC Results

125 participants enrolled in the protoMOOC, but not all of them were active learners during the course. 43 learners participated in the week 1 activity, which was about getting to know the learning community (not graded). 30 learners submitted the first assignment on *Observation* and 20 submitted the second one on *Qualitative Interviewing*. Below, we describe the results of CES, SCR and qualitative interviews.

4.1 Course Evaluation Survey

4.1.1 Pre CES

70 learners participated in the pre CES. Most of them belonged to the age group of 25–34 year-olds (n = 33). More women (n = 40) than men (n = 29) participated in the survey. Regarding learners' experience levels, 33 were new to design thinking and 30 never took a MOOC before. This allowed us to learn more about the amplified needs of beginners to both content and online learning.

4.1.2 Post CES

The post CES counted 20 learner ratings. Most of those 20 learners showed overall satisfaction with the MOOC: On a scale of 1 (not satisfied at all) to 10 (absolutely satisfied), 6 learners rated their satisfaction with the overall MOOC with a 9, followed by satisfaction ratings of 8 and 6 by 4 learners each. Learners also voiced their suggestions for improvement on some aspects of the course, including: clearer task descriptions for exercises and assignments, better communication of deadlines, more examples of design thinking application in real life, and a summary of the course content.

We also asked learners how they perceived the course requirements and assignments. 75% of them were satisfied with the workload. Learners rated assignments as the most valuable learning mode in the course, followed by exercises and videos (see Fig. 2). Finally 75% stated that their personal learning objectives were met in this course.

There are many ways to illustrate the design thinking process; the six bubbles (HPI D-School Potsdam) and the five hexagons (d.school Stanford) to name just two. Inspired by IDEO.org's presentation of the human-centered process, we decided to introduce the following three phases to represent different working modes in the design thinking process: finding inspirations, creating solutions, and validating with users. For the *Inspirations for Design* MOOC, we decided to focus

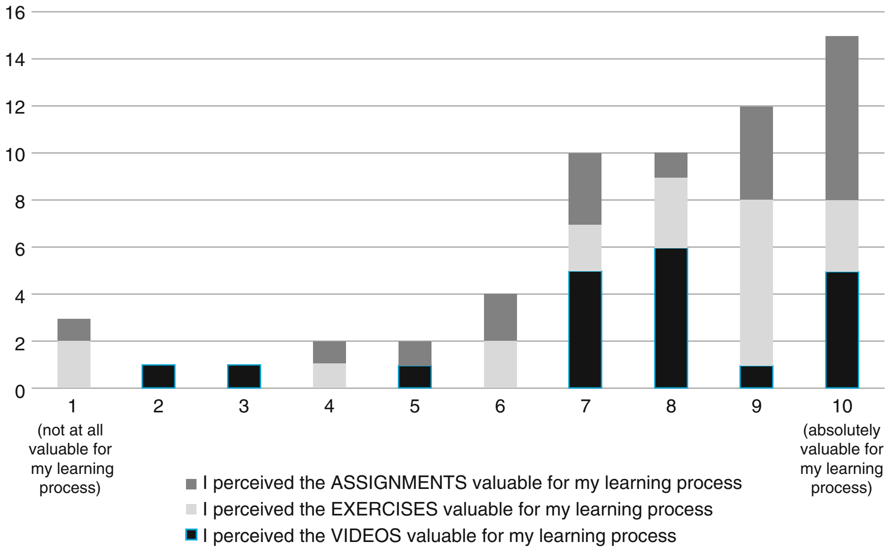


Fig. 2 Learners' perceptions of how valuable (the) different learning formats were to their learning process

on two basic but important methods in the *Empathizing* phase: *Observation* and *Qualitative Interviewing*.

The reason for our choice is twofold. It allows communicating the essence of the method in a simple way to novices and it offers a clear structure for future online learning units (two successive courses). By focusing on the inspiration phase of the design thinking process and specifically conveying major and powerful skills to detect inspirations for design, such as being attentive to one's surrounding (*Observation*) or finding insights and inferring meaning (*Qualitative Interviewing*), we explored how design thinking skills can be conveyed through online learning. Accordingly, we plan to create new and more design thinking learning units in this research project (i.e., a consecutive MOOC covering skills that are crucial during ideation and prototyping).

4.2 *Skill Confidence Rating*

We analyzed the SCR by calculating all means. 49 learners filled in the pre SCR for the first module (Observation), and 24 learners filled in the corresponding post SCR. In the second module on *Qualitative Interviewing*, n = 40 participated in the pre SCR and n = 24 in the post SCR. Means for all question items are displayed in Table 1.

Table 1 Mean comparison of SCR pre and post rating for all question items per learning module

Module	Pre	Post
Observation		
1. To what extent would you sinder yourself being attentive to your daily environment (before and after taking this week's learning unit on Observation)	M = 7.29 (n = 49)	M = 9.33 (n = 27)
2. How easy is it for you to interpret what lies behind a problem (before and after taking this week's learning unit on Observation)	M = 7.38 (n = 49)	M = 9.71 (n = 27)
Qualitative interviewing		
How confident do you feel about preparing for a qualitative interview (e.g., writing an interview scheme), (before and after taking this week's learning unit on Qualitative Interviewing)	M = 6.59 (n = 40)	M = 8.30 (n = 24)
How confident do you feel about conducting qualitative interviews (before and after taking this week's learning unit on Qualitative Interviewing)	M = 6.69 (n = 40)	M = 8.30 (n = 24)
How confident do you feel about inferring meaning from your interview results (before and after taking this week's learning unit on Qualitative Interviewing)	M = 6.77 (n = 40)	M = 8.26 (n = 24)

We found a skill confidence mean increase for both learning modules in all question items. Although the result is limited by dissimilar sample sizes, it can be assumed that the course content helped learners develop more confidence with the taught skills.

The open-ended question item in the post SCRs contained remarks and constructive feedback from learners. We categorized these comments in a so-called feedback grid, which is a commonly used tool in design thinking to collect and cluster user feedback. The feedback grid consists of four categories: positive comments (+), negative comments (−), open questions (?) and new ideas for improvement (!). Table 2 summarizes all SCR feedback. In general, learner comments were positive, but we also extracted two major needs that referred to the protoMOOC's didactic design. On the one hand, learners stated a need for additional information on topics like nonverbal communication in interviews or hidden needs to deepen their skills. On the other hand, they wished for more elaborated feedback on their performance.

4.3 *Qualitative Interviews*

We used the design thinking tool of feedback grids again to structure notes from the 16 qualitative interviews. Figure 3 shows an example of the categorization process.

Table 2 Evaluation of the open post skill confidence rating questions listed by learning module

Feedback grid category	Module: Observation	Module: Qualitative interviewing
(+) Positive comments	– Enjoyment and fun (to look out for daily workarounds)	– Well-balanced content and learning speed – Stimulating
(−) Negative comments	– Technical problems – Unclear learning objective/ weekly goal – Lacking more knowledge on how to present a job story	– Missing concrete feedback on interviewing skills
(?) Open questions	– Possibility of workarounds beyond products: "service workarounds?" – Unclear learning objective/ weekly goal – Lacking more knowledge on how to present a job story	/
(!) New idea	– Enhance poignancy of topical session segments by adding more examples – More case studies – Add more content to observation skill: watching actions	– How to do a research plan in advance exactly? Which tools could help to do this? – Add text in videos – More content (e.g., on hidden needs, motivation, non-vocal reactions, creating rapport) – Advice on interview setting – Receive feedback on uploaded interview

5 MOOC Iteration

In the following, we derive the process from feedback grids to actionable tasks for iteration. After structuring the feedback gathered through CES, SCR and the qualitative interviews in feedback grids, we derived the following 14 topics from the grids:

- videos
- templates
- additional resources
- structure, time commitment and deadline
- course title and objectives
- new ideas for course design
- research methods (survey design, etc.)
- learning module: warm-up
- learning module: *Observation*
- learning module: *Qualitative Interviewing*
- peer review assignments

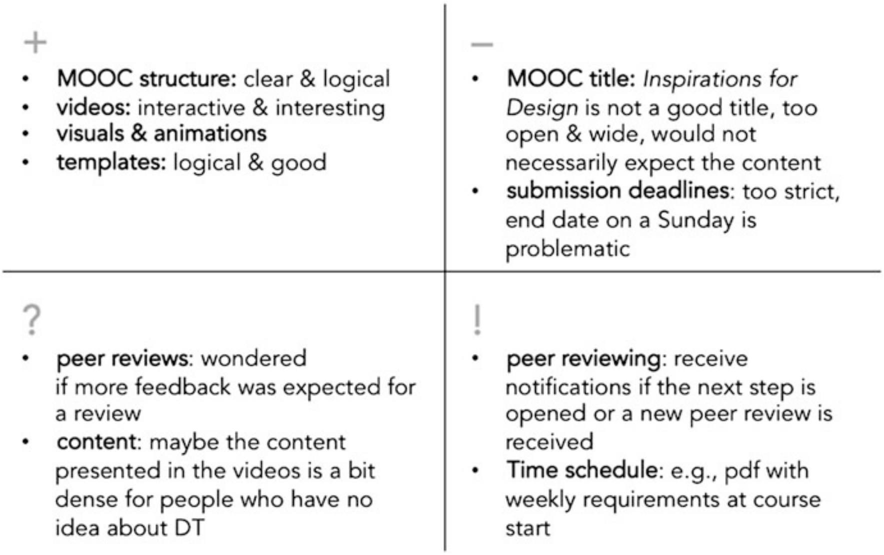


Fig. 3 Examples from clustered feedback based on 16 qualitative interviews

- discussion forum
- addressing cultural diversity
- other

We consequently merged the feedback grids of CES, SCR and qualitative interviews in one "Topical Feedback List," where comments on the 14 topics were listed below. In the next step, we translated the feedback for each topic into actionable tasks by using the "Keep/Change/Introduce" model, a modification of the "Stop/Start/Change/Continue" model for feedback management.

Figure 4 shows an example of how we translated the feedback from CES, SCR and follow up interviews for the topic of "learning module: *Qualitative Interviewing*" into an actionable iteration goal.

Learners in the protoMOOC conducted and interpreted a qualitative interview within 1 week. CES, SCR and interview findings indicated that the workload for the Q*ualitative Interviewing* module was high. We translated this into the actionable iteration task of changing the *Qualitative Interview* module length from 1 to 2 weeks by splitting up the tasks of conducting an interview and taking the first steps towards interpreting it. Consequently, we created new video lectures, templates and assignments for the iterated MOOC. In the new version, the first week (*Qualitative Interviewing*) contained videos and guidelines on how to prepare and conduct a good interview, and the second week contained videos and templates on how to interpret the interview notes. We defined 82 tasks to iterate aspects of the protoMOOC, from which we applied 57 for the first MOOC iteration. In Table 3, we show the most important changes for the 14 topics.

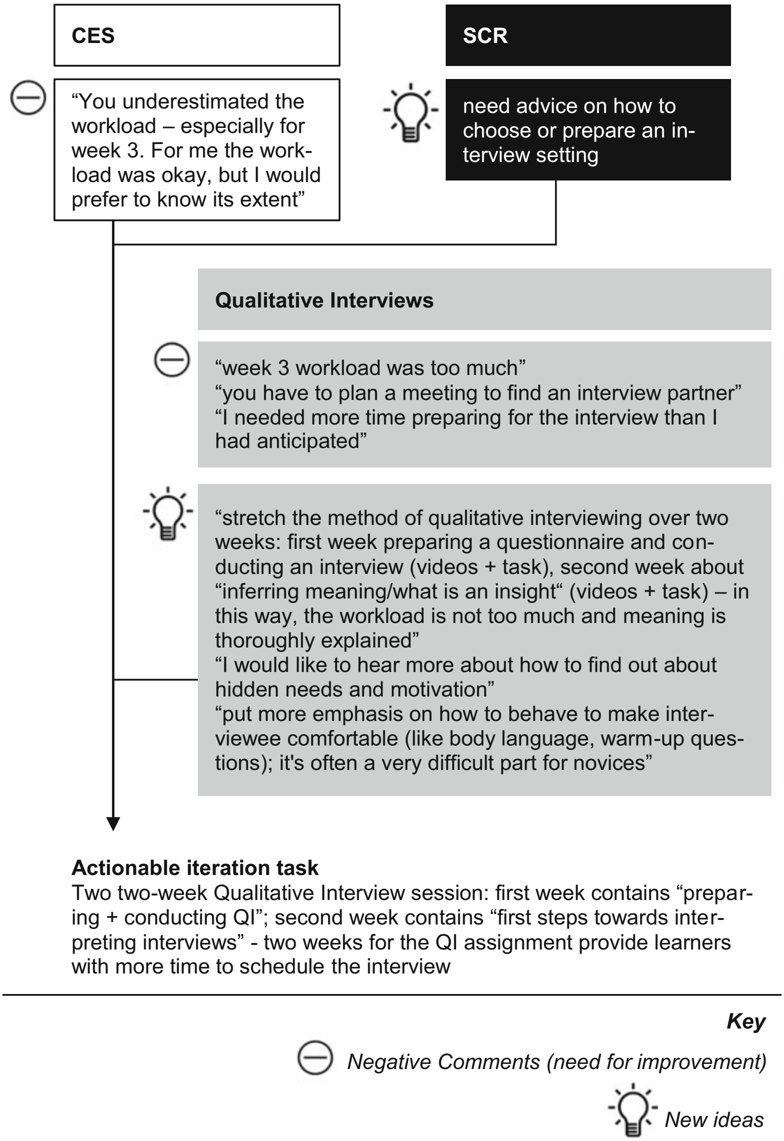


Fig. 4 Example of the iteration process for the topic “Learning Module: Qualitative Interviewing.” “Quotation marks” indicate direct learner quotes we collected from the CES, SCR and qualitative follow up interviews

Table 3 The most important iteration tasks for the 14 feedback topics derived from participant feedback

Feedback topic	Most important iteration task
Videos	Creation of seven new course videos to condense and complement content and iteration of existing videos
Templates	Iteration and creation of assignment and exercise templates
Additional resources	Expansion of additional resources, including case studies from different industries
Structure, time commitment and deadlines	Creation of clear time structure visualizations for the landing page
Course title and course	Change of course title for clarity
Course design	Creation of four expert videos to enhance the course content
Research methods (survey design etc.)	Introduction of "knowledge transfer" discussion forum posts as means of gathering information and encouraging skill application
Learning module "Warm-up"	Keep introduction week
Learning module "Observation"	Creation of additional material for contextualizing the "job story" method
Learning module "Qualitative Interviewing"	Restructuring and enhancement of unit to two weeks length
Peer review assignment	Iteration of assignment rubrics, creation of example assignments and example feedback
Discussion forum	Introduction of "Bug List" post to report about possible technical errors
Addressing cultural diversity	Adaption of assignment texts to emphasize cultural sensitivity
Other	Creation of two module summaries in textbook format

6 Mooc Iteration: First Results from the Public MOOC

The iterated public MOOC ran from August to October 2017 on the elearning platform openHPI. 4398 learners were enrolled at the beginning of the course, 5164 learners were enrolled in the middle of the course. 64% of these participants (n = 3043) were counted as "shows" in the course (meaning they visited the course at least once). Regarding assignment submissions, 932 learners submitted the first assignments, and 619 learners submitted the second assignment. Learners who reached more than 50% of total points in the course received a Record of Achievement (n = 786). Points were awarded through assignments and exercises.

The discussion forum displayed an active aspect of this MOOC. There, learners interacted with each other, raised questions and commented on various topics throughout the course. In total there were 191 discussion threads containing 1912 posts. To assess learners satisfaction in the iterated MOOC, we again inserted a CES bracketing the whole course. Learner development was assessed with SCR surveys bracketing the topical sessions.

Results from the post CES show that learners were satisfied with the overall course. As shown in Fig. 5, 84.75% of participants who answered the question item

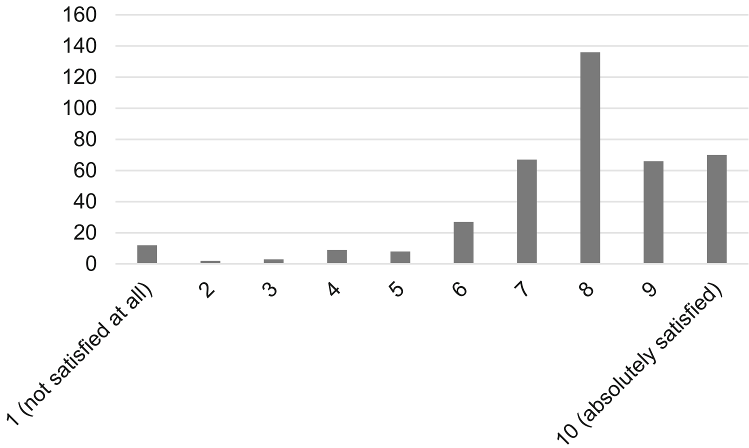


Fig. 5 Post CES results for participants' rating of satisfaction with the overall course (n = 400)

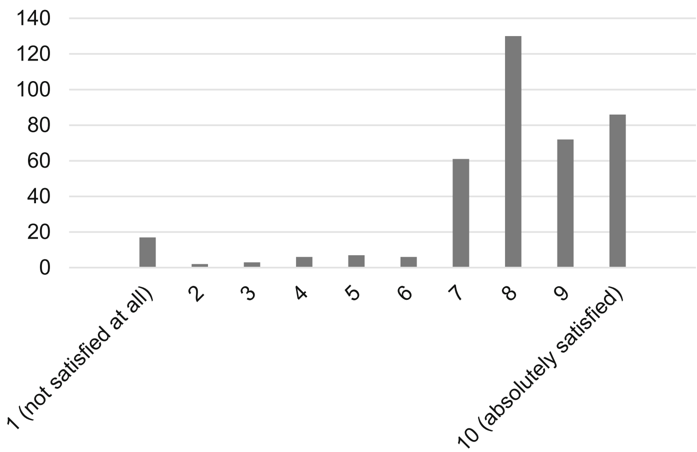


Fig. 6 Post CES results for participants' rating of satisfaction with the quality of the course content (n = 390)

(n = 400) chose a value between 7 and 10 on a scale from 1 (not satisfied at all) to 10 (absolutely satisfied). 136 participants rated their satisfaction with 8, followed by satisfaction ratings of 10 (by 70 participants) and 7 (by 67 participants).

We see a similar result in the responses of the question item "How satisfied are you with the quality of the content presented in the course?". 89.49% of participants who answered the question item (n = 390) chose a value between 7 and 10 (see Fig. 6).

Some learners voiced difficulties with managing their time for the weeks of conducting and interpreting qualitative interviews in the survey and stated that they got "stressed out" during week 3 and 4. Nonetheless, 84.89% rated the weekly

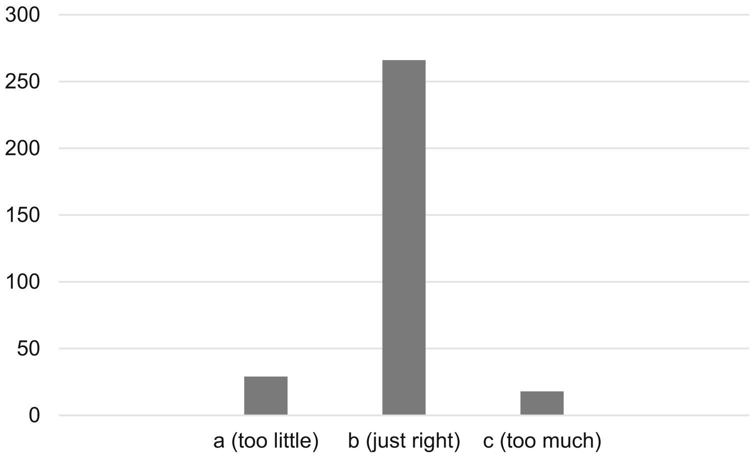


Fig. 7 Post CES results for participants' rating of the weekly workload (n = 313)

workload of the course as "just right"—266 participants out of n = 313. See Fig. 7 for detailed results.

7 Conclusion and Outlook

Overall the iterative approach to MOOC design proved to be effective and positive. Feedback gathered from the protoMOOC led to 57 improvements. Testing the course in a closed learner group helped to detect and deal with potential technical errors before scaling it to a large international learner crowd. CES and SCR results show an overall learner satisfaction with the iterated MOOC. Participant numbers were steady, with the number of submissions decreasing from the first (n = 932) to the second assignment (n = 619). The team aims to apply their learning in MOOC design to create successive MOOCs and learning experiences on other aspects of design thinking.

In the future, we want to follow up on our initial attempts to make the MOOC more adaptive while it is running. One example of efforts towards adaptability are "wrap up" and "on-the-go" videos in which we clarified repeated questions from the discussion forum, shared our observations during the course (learner activities) with the participants and showcased top-rated learner assignments. The feedback shows that learners appreciated these course elements. Another example for the adaptive role we took as instructors was to make minor changes to the course structure based on learner comments (e.g., providing all assignment templates for download in the last course week).

In the future, we intend to deepen and expand adaptive elements which we introduced and tested in this iterated MOOC.

Acknowledgements We thank the HPI-Stanford Design Thinking Program for enabling this project. For the MOOC production, we relied on the advice of Stefanie Schweiger and Thomas Staubitz (openHPI) and the technical support of the HPI Medientechnik team. Many thanks to Dr. Sharon Nemeth for copyediting and language support.

References

Bloom, B. S. (1994). In L. W. Anderson & L. A. Sosniak (Eds.)., Bloom's taxonomy: A forty-year retrospective *Reflections on the development and use of the taxonomy* (pp. 1–8). Chicago: University of Chicago Press.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *AAHE Bulletin, 3*, 3–7.

Kraiger, K., Ford, J., & Salas, E. (1993). Application of cognitive, skill-based and affective theories of learning outcomes to new methods of training education. *Journal of Applied Psychology, 78* (2), 311–328.

Siemens, G., & Tittenberger, P. (2009). *Handbook of emerging technologies for learning*. Manitoba: University of Manitoba.

Staubitz, T., Petrick, D., Bauer, M., Renz, J., & Meinel, C. (2016, April). Improving the peer assessment experience on MOOC platforms. In *Proceedings of the third (2016) ACM conference on Learning@Scale* (pp. 389–398). ACM.

Taheri, M., & Meinel, C. (2015). Pedagogical evaluation of the design thinking MOOCs. In *Proceedings from the 3rd international conference for design education researchers* (pp. 469–481). Chicago: Design Research Society.

Taheri, M., Mayer, L., von Schmieden, K., & Meinel, C. (2017). The DT MOOC prototype: Towards teaching design thinking at scale. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research* (pp. 217–237). Berlin: Springer.

Crowd Research: Open and Scalable University Laboratories

Rajan Vaish, Snehalkumar (Neil) S. Gaikwad, Geza Kovacs, Andreas Veit, Ranjay Krishna, Imanol Arrieta Ibarra, Camelia Simoiu, Michael Wilber, Serge Belongie, Sharad C. Goel, James Davis, and Michael S. Bernstein

Abstract Research experiences today are limited to a privileged few at select universities. Providing open access to research experiences would enable global upward mobility and increased diversity in the scientific workforce. How can we coordinate a crowd of diverse volunteers on open-ended research? How could a PI have enough visibility into each person's contributions to recommend them for further study? We present Crowd Research, a crowdsourcing technique that coordinates open-ended research through an iterative cycle of open contribution, synchronous collaboration, and peer assessment. To aid upward mobility and recognize contributions in publications, we introduce a decentralized credit system: participants allocate credits to each other, which a graph centrality algorithm translates into a collectively-created author order. Over 1500 people from 62 countries have participated, 74% from institutions with low access to research. Over 2 years and three projects, this crowd has produced articles at top-tier Computer Science venues, and participants have gone on to leading graduate programs.

1 Introduction

Scientific research remains the domain of the privileged few. Those blessed with the socioeconomic opportunity to attend prestigious universities can gain access to research experiences that support open-ended inquiry, train scientific minds and

R. Vaish (✉) · G. Kovacs · R. Krishna · I. A. Ibarra · C. Simoiu · S. C. Goel · M. S. Bernstein
Stanford University, Stanford, CA, USA
e-mail: rvaish@cs.stanford.edu; msb@cs.stanford.edu

S. S. Gaikwad
MIT Media Lab, Cambridge, MA, USA

A. Veit · M. Wilber · S. Belongie
Cornell Tech, New York, NY, USA

J. Davis
UC Santa Cruz, Santa Cruz, CA, USA

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_7

Fig. 1 We present a crowdsourcing technique that has enabled access to research experiences for over 1500 people from 62 countries. Participants achieve upward educational mobility while creating research systems and co-authoring papers at top-tier ACM venue

launch careers (Russell et al. 2007). Unfortunately, these opportunities remain out of reach for the vast majority of people worldwide (Bianchini 2011; Bowen and Bok 2016; Page 2008). Such people may have the creativity, insight, and work ethic to produce major achievements, but lack access to the opportunity. The result is an ecosystem that systematically underrepresents minorities and developing regions, and a literature that overlooks their diverse perspectives.

Providing open access to research experiences would open new channels for upward educational and career mobility worldwide. However, how can a principal investigator such as a faculty member or research scientist coordinate an entire crowd of diverse people? If the goal is to give participants full breadth to demonstrate creativity, solve unanticipated challenges, and guide the project's direction—not to reduce them to mechanical research assistants—no general techniques yet exist. Citizen science efforts have pursued protein folding (Cooper et al. 2010; Lee et al. 2014), scientific dataset labeling (Land et al. 2008; Sullivan et al. 2009), math proofs (Cranshaw and Kittur 2011), and experiment replication (Open Science Collaboration 2015), but these projects required one pre-defined, static goals rather than allowing participants to iteratively guide the research goal. In addition, sheer scale prevents a principal investigator from having full visibility into each participant's contributions, threatening their ability to recommend participants for further study.

This chapter describes *Crowd Research*, a crowdsourcing technique that enables open access for a global crowd to work together on research under a principal investigator (PI) (Fig. 1). Crowd Research participants collaborate online as one large team to brainstorm research ideas, execute solutions, and publish scholarly articles—a university laboratory at massive scale. To facilitate open access, we introduce a crowdsourcing technique that comprises weekly cycles of contribution, synchronous collaboration, and peer assessment to produce each next week's iterative goal. A suite of systems carries research from ideation to execution, including brainstorming, engineering, design, analysis and paper writing. For the PI, Crowd Research offers a chance to convene hundreds or thousands of people on a single massive project, enabling research achievements at a scale that is rare today.

To enable upward career and educational mobility, Crowd Research must provide contributors with credible evidence of their impact. However, a PI cannot easily disaggregate participants' interdependent contributions, and may not have centralized visibility into each participant's work. We thus introduce a decentralized credit system where participants allocate credit to each other. This allocation process creates a weighted directed graph, enabling our system to apply a graph centrality algorithm to determine a collectively-created author order for publication and the PI's recommendation letters.

Crowd Research has so far brought together over 1500 participants from six continents, 74% of whom come from universities ranked below 500 in global research activity and influence by Times Higher Education (2017). It has included three different research projects with four PIs from Stanford, UC Santa Cruz and Cornell—ranging from human-computer interaction (HCI) to data science to artificial intelligence (AI). These projects produced crowd-authored papers that have been accepted to top-tier Computer Science venues including ACM UIST (Gaikwad et al. 2016) and ACM CSCW (Whiting et al. 2017). Participants have leveraged their contributions to receive recommendation letters from PIs. Despite having a median of zero other letter writers from institutions ranked above 500 worldwide, participants have been admitted for further study at undergraduate and graduate programs at universities such as Stanford, UC Berkeley, Carnegie Mellon, and MIT.

Our contributions span a crowdsourcing technique for coordination of open-ended, long-term and complex goals; a decentralized method for allocating credit; and an analysis of a 2 year, large-scale deployment of the method. To follow, we position Crowd Research in related work, describe the technique, and analyze our deployment and limitations.

2 Related Work

Research experiences are not just authentic practice: they impact upward mobility. Engaging in research increases the probability of enrolling in STEM graduate programs (Kevin Eagan et al. 2013), both for professional degrees (Lopatto 2004) and PhDs (Zydney et al. 2002). Research experiences also increase interest in STEM careers (Russell et al. 2007) and increase a student's likelihood of using faculty recommendations for jobs (Hathaway et al. 2002). Other improvements accrue to technical skills, interpersonal skills, and scientific literacy (Cronje et al. 2011; Eric Landrum and Nelsen 2002).

The size of Crowd Research and the diversity of its membership offers new opportunities for science and engineering research. Having many people brings a diversity of ideas (Lakhani et al. 2013; Mao et al. 2016; Lixiu et al. 2014). Diversity arguably brings even greater benefits: diverse problem solvers outperform groups of high-ability problem solvers (Hong and Page 2004), and diverse perspectives unearth hidden assumptions and yield more active and effortful thought (Gurin et al. 2002). Our work synthesizes these benefits by drawing on a diverse worldwide population, and applying them toward open-ended research goals. In doing so, it trades off the expertise that most research projects can assume of their participants (e.g., graduate level coursework), and instead uses peer assessment (Kulkarni et al. 2013), the web, and lectures to provide on-demand training.

Research experiences have traditionally been one-on-one cognitive apprenticeships (Hunter et al. 2007). Providing mentorship is a nontrivial time commitment for faculty, which often limits who can be mentored (Zhang et al. 2017). More critically, universities which produce the world's most-cited research are concentrated in North America and Europe (The Times Higher Education World University Rankings 2017), far from the world's largest and developing population centers. Crowd Research introduces techniques to bring the benefits of research experiences to a far larger group.

2.1 Online Access to Training and Science

Crowd Research draws lessons from online education and citizen science, each of which expands access to opportunities that are typically only available within universities. MOOCs democratize access to online learning opportunities (Dillahunt et al. 2014), offering an attractive template for Crowd Research. Unfortunately, MOOCs are especially likely to leave behind people from less developed areas (Kizilcec and Halawa 2015), and taking a MOOC does not translate to upward career mobility (Dillahunt et al. 2016). Crowd Research builds on these efforts by directly encouraging authentic practice and enabling calibrated assessment for upward mobility.

Citizen science enables members of the public to contribute to research (Silvertown 2009). Typically, these projects predefine the goal of the project and the method of contribution, and participants contribute by filling out the "rows" of the desired dataset. For example, projects engage volunteers to upload bird locations on eBird (Sullivan et al. 2009), take tests on LabInTheWild (Reinecke and Gajos 2015), and label galaxies on Zooniverse (Cox et al. 2015). Other projects give participants more freedom in how they answer a research question, for example, crowdsourced math proofs in the Polymath project (Cranshaw and Kittur 2011) and protein folding in Foldit (Khatib et al. 2011). Crowd Research represents a rarer third category, co-created projects, where participants are involved not just in data collection and execution but also in the conception and ongoing evolution of the research (Bonney et al. 2009; Oliveira et al. 2017; Pandey et al. 2017). Crowd Research is unusual even in this class of projects because participants own the whole research arc, rather than one focused part.

While citizen science has succeeded in engaging with participants worldwide, it has struggled to close the access gap. For example: (1) Zooniverse participants tend to be from highly educated countries (Jordan Raddick et al. 2009); (2) Nearly all Polymath participants were faculty or Ph.D. students, 86% had published papers, and only one was known to be female (Cranshaw and Kittur 2011); and (3) OpenStreetMap contributors are 96% male, with three-quarters holding a post-graduate degree (Budhathoki 2016). However, some citizen science projects have explicitly attempted to incorporate marginalized communities (Stevens et al. 2014). Crowd Research reaches a global audience via diversified recruiting and provides direct incentives for upward mobility such as paper authorship and recommendation letters. However, Crowd Research cannot yet overcome internet infrastructure and language limitations.

2.2 *Coordinating Research and Crowds*

To enable crowds to engage in collaborative research, Crowd Research extends work from CSCW and social computing. Authentic tasks (e.g., Suzuki et al. 2016) and feedback (e.g., Dow et al. 2012; Kulkarni et al. 2015) are both critical elements to improvement. Within traditional laboratory environments, pair working sessions can support knowledge transfer (Miller et al. 2014), and agile research studios can scale mentorship per PI to about 20 students (Zhang et al. 2017). Crowd Research operates at a much larger scale and with more diverse participants. This requires different approaches, in particular fewer team-based agile methods and more structured, pre-defined milestones. Decision making must also become more decentralized, e.g., via peer assessment, decentralized credit allocation, and DRIs.

Crowdsourcing techniques increasingly aim to support complex outcomes (Kittur et al. 2011, 2013). These systems can now support goals ranging from software engineering (Chen et al. 2016; LaToza et al. 2014) to writing (Nebeling et al. 2016; Teevan et al. 2016). Crowd Research shares some characteristics with this work, organizing the crowd into expertise-based teams (Retelny et al. 2014) and hierarchical structures (Valentine et al. 2017) that can adapt as the crowd proceeds. Unlike prior work, Crowd Research is designed to train participants, so it introduces explicit peer feedback and direct engagement with the PI as a leader.

2.3 *Determining Credit*

To aid upward mobility, Crowd Research must provide assessments of participants' contributions that they can leverage for school and job applications. One strategy for determining author order is to alphabetize. However, women receive less credit than men in alphabetical author orders (Sarsons 2015). A second strategy is to publish as a single joint author, as in "DHJ Polymath" in the Polymath Project (Gowers and Nielsen 2009). However, a joint name does not provide strong signals for participants to leverage for recommendations. Firms such as Quirky and Assembly offer credit for pre-defined contribution categories, e.g., 1% for coming up with the product's name. However, research is an iterative process where it is not always clear which contributions will wind up being influential. So, we develop a new, decentralized credit approach.

Prior work has studied algorithmic ranking schemes, for example hubs and authorities (Kleinberg 1999) and PageRank (Page et al. 1999). Similar schemes have been applied to curation (Haizi et al. 2016) and citations in order to determine influence (Ding et al. 2009; Shen and Barabási 2014; Sun et al. 2009). However, these approaches all assume the existence of a network, which Crowd Research does not have. So, Crowd Research introduces a technique that allows all participants to have a say in the eventual allocation of credit, translating credit into a graph problem.

3 Crowd Research

Crowd Research (Fig. 2) introduces a crowdsourcing technique to enable worldwide access to research experiences without overwhelming a PI. In this section we present the approach in detail, oriented around (1) how Crowd Research coordinates large groups of participants, (2) what systems enable collaboration and scholarly outcomes, and (3) how it enables upward mobility through a decentralized credit system.

3.1 Coordination Strategy and Process

Crowd Research enables thousands of people online to coordinate joint progress on an open-ended research effort. Prior work has often pre-structured the crowd's contributions—for example providing interfaces for folding proteins (Khatib et al. 2011)—because the goal and the tools needed for success could be de-fined a priori. Many researchers have eschewed crowdsourcing for exactly this reason: "the process of discovery can be highly uncertain, iterative, and often serendipitous", making the reduction to a crowdsourcing process "hard to imagine" (Law et al. 2017). So,

Fig. 2 Crowd Research comprises weekly meetings to discuss the project, milestones to submit concrete progress, and peer assessment to identify top submissions

Crowd Research introduces an iterative crowdsourcing technique based on milestones and peer assessment that allows the effort to iterate and adapt over time.

We will refer to the roles of *PI (principal investigator)*, who advises the project; *RA (research assistant)*, who supports logistics, and *participants*, who are members of the crowd. The PIs' motivation was to tap into a diversity of perspectives, mentor far more students than they normally could in their careers, and try out more ambitious projects than typical in their labs. Each PI recruited two RAs to help. The RAs put in a few hours per week—in no cases were these projects the RAs' primary research—mainly helping with onboarding, analyzing top submissions, and answering logistical questions.

3.1.1 Open Call Recruitment

The first step in Crowd Research is to recruit a crowd. Crowd Research opens with a global online call inviting people to join one of the available posted projects. A public web page describes the opportunity, the PIs involved, and their institutions. We shared this page via social media on Twitter and Facebook groups, cold emails to faculty at international universities, and publicly accessible mailing lists. Interested participants have several weeks to sign up alone or in teams. While selective recruitment is possible, to maximize accessibility we accept all participants who signed up, and create accounts for them on our collaboration platforms.

We launched three different Crowd Research projects, helping us understand how Crowd Research differs across different PIs and research areas. The PIs chose and seeded initial ideas for the projects, much like an initial idea might be seeded with a traditional graduate student. Each PI later developed the idea in collaboration with the crowd.

First, the human-computer interaction (HCI) project, led by Prof. Michael Bernstein at Stanford University, set out to create a new paid crowdsourcing marketplace a la Amazon Mechanical Turk. In current crowdsourcing marketplaces, workers feel disrespected, and requesters do not trust the results they receive (Irani and Silberman 2013; Martin et al. 2014). The HCI project works on designing, engineering, and studying a new crowdsourcing marketplace, Daemo, to improve work quality and give workers governance of the platform. Second, the computer vision project, led by Prof. James Davis at UC Santa Cruz and Prof. Serge Belongie at Cornell Tech, seeks to improve visual classification accuracy. Integrating off-the-shelf machine classifiers with paid crowds is challenging (Russakovsky et al. 2015). This project explores strategies to increase accuracy and decrease cost under this setting. Third, the data science project, led by Prof. Sharad Goel at Stanford University, seeks to design and run the world's largest "wisdom of crowds" experiment. There is still little consensus on how generally the wisdom of crowds phenomenon holds, how best to aggregate judgments, and how social influence affects estimates. This project tests these boundary conditions by collectively designing and developing 1000 different prediction tasks in 50 subject domains, and launching them as a large-scale meta-experiment.

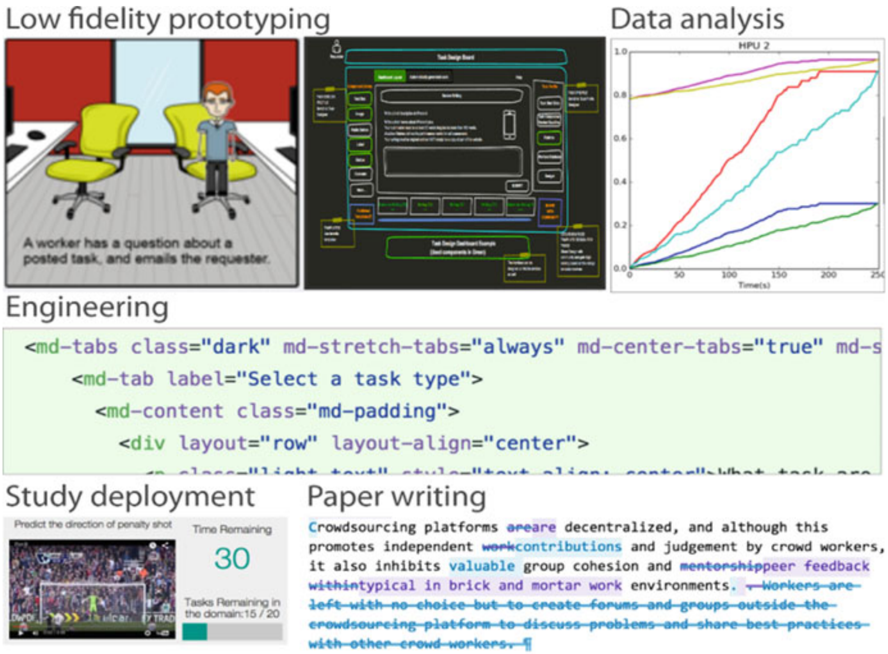


Fig. 3 Milestones included prototyping, engineering, and writing

3.1.2 Milestone Submission

Each week, the PI identifies a concrete goal for the crowd, called a *milestone*. Milestones are scoped at 5 to 10 hours of work per week. Past milestones have included (1) participating in a needfinding interview with Mechanical Turk users, (2) engineering an experimental scaffold for experiments, (3) proposing experimental designs, (4) implementing a proposed algorithm from a previous week, and (5) brainstorming iterations of the research idea based on feedback from the PI (Fig. 3). For example, one milestone in the HCI project was focused on needfinding, and involved reading papers, joining a panel interview with workers and requesters, and then synthesizing insights. Each project maintains a wiki where the PI or an RA uploads details for all milestones.

Participants work in parallel during the week, individually or in teams, to complete the milestone. Since the process works on a weekly cycle, participants have about 6 days to complete each milestone. The Slack group chat platform operates as a brainstorming and discussion room for participants where they can interact with each other, help each other, and ask questions. Slack can become quite busy. To manage it, some channels (e.g., #announcements) are low traffic and intended to be read in their entirety. Others are busy and scoped narrowly to a milestone area (e.g., #design, #engineering). Participants often create ad-hoc channels for each milestone and other interest-based channels (e.g., #highschoolers,

#machinelearning) to meet other like-minded participants. This helped participants selectively follow relevant conversations without getting overwhelmed.

In the early phases of the project, milestones were limited to one goal each week. However, it soon became clear that the crowd brought many different skills to the projects, and some participants would wait for weeks for their skills to be applicable. So, we began to allow multiple parallel milestones each week, enabling participants to self-organize and select which ones to complete. For example, 1 week's milestones might include creating interaction mockups for designers, a front-end feature implementation for AngularJS engineers, and a back-end feature implementation Django/Python engineers.

At the end of the week, teams submit their milestones to a peer assessment system. Participants create a page on the wiki containing their milestone submission, and submit that link.

3.1.3 Peer Assessment

At this point, there are a large number of submissions to the milestone—far too many for the PI to read and synthesize. They vary greatly in quality, content, and coherence. The next stage of Crowd Research harnesses peer assessment to give feedback on the submissions and provide a rough ranking so that the PI can concentrate on the most promising ideas. The peer assessment process is open for 1 day, with a cutoff for feedback a few hours before the weekly team meeting.

Our peer assessment system functions similarly to a social aggregator such as Reddit. Once the submission deadline has passed, the crowd can look at each others' submissions, leave comments and upvote strong submissions. The PI and RAs choose a default sort for the system: e.g., most upvotes at the top to encourage feedback on promising candidates, or fewest comments at the top to encourage diversity. While we initially experimented with a system that randomized a double-blind assignment, anonymous feedback was needlessly negative and evaluative. Instead, we shifted to a system (Fig. 4) where participants' names, submissions and feedback were all public, participants chose to give feedback to, prompting a more positive environment.

At the conclusion of the feedback period, the process has called out some of the most interesting and inspirational submissions. This set is small enough for the PI or RAs to collate. They read these submissions and use them as the basis for discussion in the weekly team meeting. While an upvoting process is not perfect (Gilbert 2013), it succeeds at separating the insightful submissions from the submissions that are ill-formed, incomplete, or do not display enough understanding.

3.1.4 Weekly Video Meeting

Once the PI and RAs read the filtered submissions, they discuss next steps with the crowd, much like a PI would with a traditional graduate student. Crowd Research

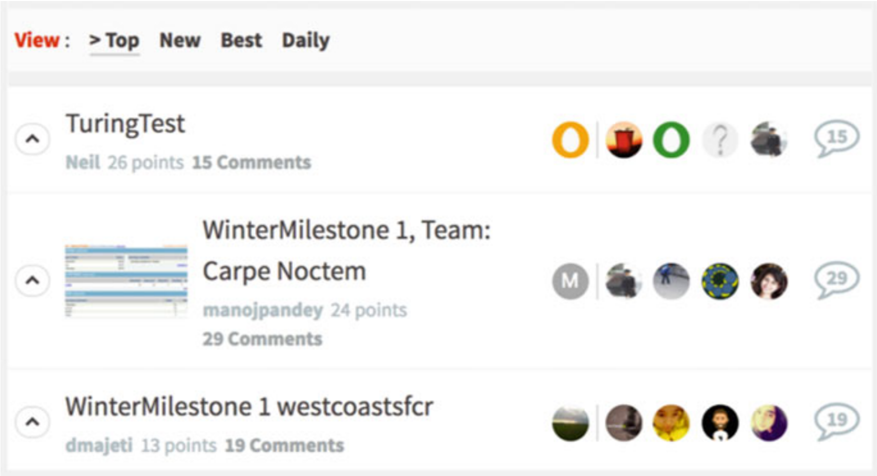


Fig. 4 Participants view each other's milestone submissions, leave comments, and upvote promising ideas

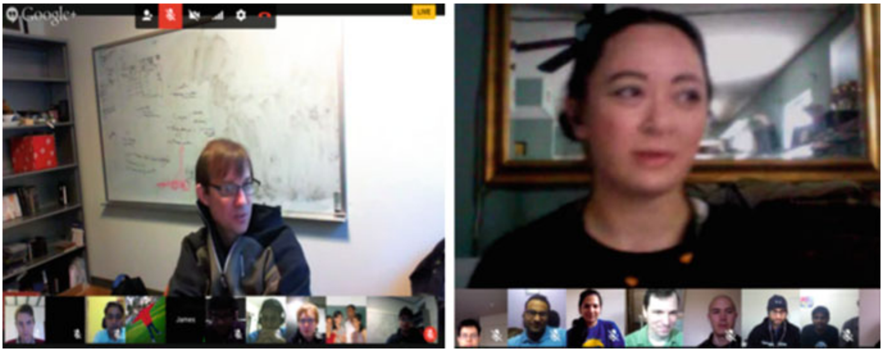


Fig. 5 Weekly video meetings on YouTube Live include participants who submitted highly-rated milestones

concludes its weekly cycle with a live video meeting broadcast via YouTube Live (Google Hangouts on Air).

This 1-hour video meeting (Fig. 5) is scheduled so that as many participants as possible can attend: morning in North America enables Europe, China and India to join. The call is streamed and archived automatically on YouTube for anyone who cannot join. The PI or RAs invite participants with highly rated submissions to join the call live and explain their ideas to the rest of the crowd and the PI. Since the group video meeting has a maximum capacity, other participants join the stream and contribute via a #meetings channel on Slack. The PI informally rotates the invitations to join the call each week to ensure a distribution of nationalities, genders, and backgrounds.

The video call re-aligns all participants, whose ideas may have diverged in many directions during the week. First, the PI begins with a *Rewind* that recaps the last week's goal and progress in case participants missed a week. Second, the PI or RAs share a synthesis of that week's highly-rated milestone submissions. Participants live on the call explain their submissions, and other participants contribute via Slack, which the PI echoes into the live call. The result of this process is that all participants, even those without highlighted submissions, reset their understanding to the "argmax" of the best work.

The PI uses the last few minutes to lay out the next week's milestone, which goes live on the wiki after the call. Participants then begin working, and the process repeats.

3.1.5 Leadership, Training, Publishing

Complementing the weekly process, we developed training and leadership structures to help focus the crowd's efforts.

3.1.6 DRIs and Ad-Hoc Teams

In the early weeks of the project, divergent ideation is essential for brainstorming research ideas, proposing algorithms and experimental designs, generating design mockups, piloting software or studies, and initial writing. However, some efforts require convergence and collective execution. For example, engineering a feature, making decisions on many different proposed directions, and writing a paper all require that participants work interdependently and collaboratively.

For interdependent milestone goals, the PI empowers a Directly Responsible Individual, or DRI (Lashinsky 2012; Retelny et al. 2014). DRIs either self-nominated or were nominated by the PI to lead a milestone based on consistently high-rated milestone submissions. DRIs take charge of a milestone for that week, coordinating any participants who want to contribute to that milestone. They organize ad-hoc video meetings, delegate, and make decisions, summarizing the results in a team submission for the milestone. Being a DRI is a recognition of a participant's contributions, empowering them to have more control over decision-making—and scaling the coordination process. Over time, DRIs overtook many of the RAs' responsibilities, and the process became more community driven.

3.1.7 Training and Enrichment

Participants do not all enter the project with sufficient knowledge of the domain. PIs have two main routes for training participants: milestones and video meeting lectures. First, with milestones, a PI can ask participants to read papers and submit commentaries, much like a traditional graduate course, to ensure that participants have the research grounding. Likewise, a milestone might include completing a coding tutorial, or participating in an experiment in order to understand how to design one. Second, with video meeting lectures, the PI can reappropriate an overview lecture from an offline class to teach the crowd a concept that will be important for the research. For example, the PI might give a lecture on one style of computer vision algorithms.

Crowd research also offers an opportunity to connect participants with famous researchers who can serve as inspirational role models. These video meetings so far have included computer scientists such as Andrew Ng (Professor at Stanford and Co-Founder of Coursera), Peter Norvig (Google Research), and Anant Agarwal (MIT and EdX).

3.1.8 Paper Writing

Massively collaborative paper writing (Tomlinson et al. 2012) requires that the crowd integrate its work into academic prose. By this phase of the project, typically a set of DRIs have arisen who can five lead writing of sections of the paper. The PI identifies model papers whose argument structure are similar to the envisioned paper, and then the crowd begins weekly writing iterations on the introduction and framing of the paper. The writing itself happens via collaborative editors. In initial paper-writing efforts, participants were hesitant to overwrite each others' prose. Transitioning to a platform that supported commenting and tracking changes (e.g., Google Docs) was key in making participants feel comfortable contributing. The PI gives feedback as they would on a student's paper. Eventually, the group submits their paper for publication.

4 Designing for a Decentralized Credit System

For Crowd Research to deliver on its promise of upward mobility, it must generate credible signals of participants' level of contribution. Participants go on to apply to graduate schools and jobs, and request recommendation letters from the PI. When applicants come from traditionally under-represented areas, the PI's recommendation letter may be the only personal assessment that the company or admission committee trusts. So it is critical for the PI to be able to specify clearly: did a given participant act in a support role, or did they take a leading role in driving the project? However, with interdependent work on open-ended research, the PI may not have visibility into everyone's contributions, and it may be challenging to disaggregate them a priori (Fig. 6).

Typical solutions to credit assignment in research and practice are centralized: they rely on a single supervisor, or a small number of peers on the team, to make the assessment. For example, the lead researcher often determines author order for all collaborators on a paper, and a worker's supervisor determines the performance review. However, no single person can have a full view of another's contributions

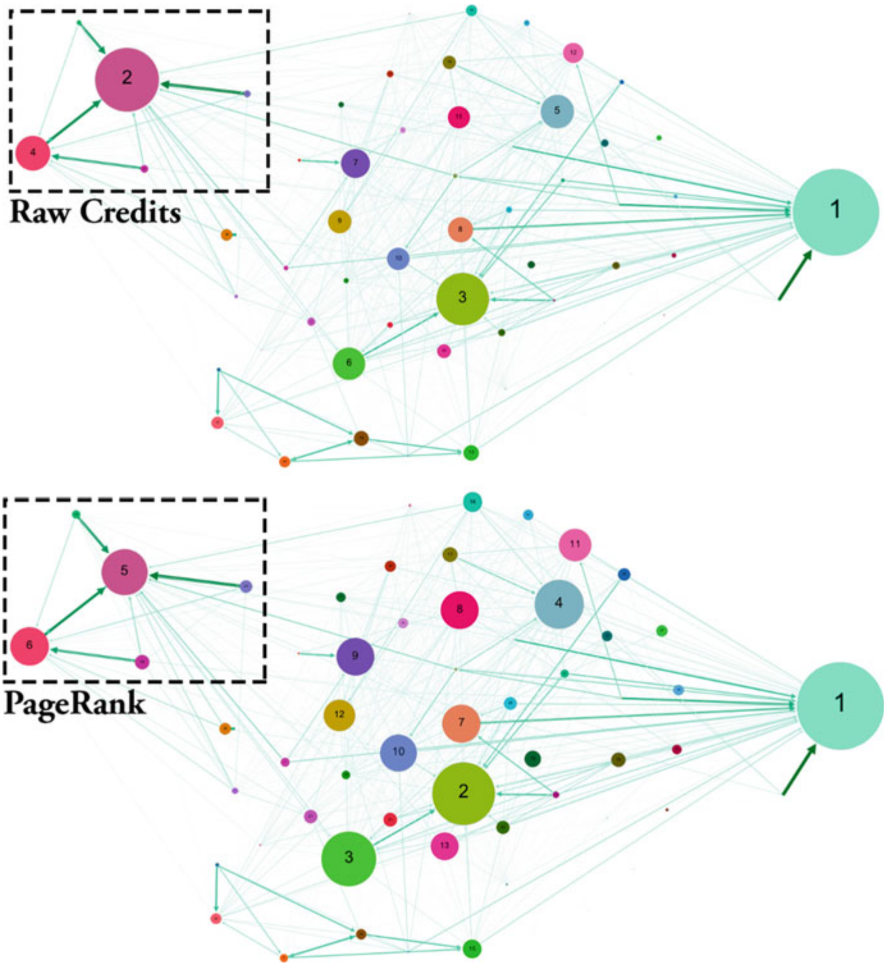


Fig. 6 The credit network from a paper submission. Edge widths indicate the number of credits given, and node diameter represents the credit score. Top: raw credits, with a link ring (top left) directing credit inward. Bottom: PageRank-adjusted credits dampen the link ring, shifting the main beneficiary from 2nd to 5th author

(Gerstner and Day 1997). So, not only can centralized credit assignment not scale to Crowd Research, but the PI would be an inaccurate assessor for many participants.

In this chapter, we introduce a *decentralized* credit system, which considers every participant's opinion in determining credit. To create a decentralized credit system, we transform the credit problem into a graph problem. This transformation allows us to draw on the tools of network science. In this approach, all participants provide peer assessments about others they have interacted with, and the algorithm aggregates these assessments to determine a final evaluation.

However, with any credit system, it is important to consider possible strategic behavior to influence author order. Not all strategies are malicious: some participants only interacted with a small percentage of the crowd. Graph centrality algorithms can help correct for these strategies. The most common form of manipulation, concentrating all credit within a small subgroup, is similar to a link ring or affiliate networks in web search. Another attack is to strategically direct credit toward others who are likely to send credit back to you—a quid-proquo strategy seen in 360-degree reviews (Toegel and Conger 2003). Our strategy must compensate for these behaviors.

In our approach, we give each participant 100 credit points that they can privately allocate to other participants based on their assessment of who impacted the project. For example, participants might assign credit to those submitting strong milestones, collaborating actively, or DRI-ing. These credit allocations create a weighted directed graph, where each node is a participant and the edge weight is the number of credits that one participant assigned to another. Intuitively, the graph encodes the credit that participants grant to each other.

We then translate the decentralized credit graph into a credit score for each participant. There are many possible transformations: we use graph centrality via PageRank (Page et al. 1999), because centrality captures the concept of a universally-recognized participant. Whereas PageRank propagates score equally across all outgoing links, we modify the algorithm to propagate scores in proportion to the outgoing edge weights to capture participants' exact credit distributions. Suppose that $G = (P,C)$ is the credit graph, where P is the set of participants and C is the set of directed weighted edges. Suppose further that $C(i, j) \in [0,1]$ represents the proportion of credits that participant i gave to j, and d is the PageRank damping factor (typically $d = 0.85$). Then our modified PageRank score $\rho_i(t)$ for participant i each iteration t is given by:

$$\rho_i(t) = \frac{1-d}{|P|} + d\sum\nolimits_{p \epsilon P} \left(\rho_p(t-1).C(p,i)\right)$$

These PageRank scores induce a ranking on participants. Given these PageRank scores, the principal investigator and DRIs work together to set a threshold score for co-authorship, based on their assessment of the level of contribution appropriate to be listed as a coauthor. Those below the cutoff are credited in acknowledgments. The PI also uses these credit rankings as a quantitative measure in letters of recommendation sent in support of participants.

We overcome the quid-pro-quo attack and link rings by manipulating the PageRank damping factor d and by limiting the fraction of a node's score that it can pass to any individual outlink (Baeza-Yates et al. 2007). However, there remain several degrees of freedom in this credit system. First: is the PI included in the credit graph? Excluding is appealing but in practice led to situations where the PI had no power to help resolve credit infighting, so we now include the PI in the credit graph. Second: when can participants see the results? In order to prevent post-submission authorship surprises, we collect initial credit distributions 1 week before the paper deadline and publicly publish a set of tentative PageRank scores. We then allow participants to change their credit distributions until a few hours before the deadline. Late credit changes affected mainly the ranking of the last authors.

5 Deployment

Evaluation of Crowd Research requires understanding (1) whether the crowdsourcing technique enabled the achievement of crowd-led research, (2) whether the technique supported access for those without traditional avenues for doing research, and (3) what impact the decentralized credit distribution technique had on contributors' rankings.

We have run three Crowd Research projects over 2 years. The projects enrolled 1697 participants from 62 countries and 6 continents, and produced crowd-authored papers at toptier Computer Science venues including ACM UIST (Gaikwad et al. 2016) and ACM CSCW (Whiting et al. 2017). Despite having a median of zero other letter writers from institutions ranked above 500 worldwide, Crowd Research participants have gone on to further study at undergraduate and graduate programs at universities such as Stanford, UC Berkeley, Carnegie Mellon University, and MIT.

5.1 Project Case Study Summaries

Participants' highest or in-progress degree was 2% high school, 73% undergraduate, 22% master's, and 3% Ph.D. 28% of participants were women, though this number varied by project: the computer vision project was overwhelmingly male, but the HCI project was 47% women. The median age was 21. Seventy one percent reported an engineering area of study. Participants included not just students and researchers but also, e.g., a data scientist on Wall Street, an ITP-trained designer, a TR35 India winner, and several professional software engineers.

5.1.1 Computer Vision: Hybrid Vision Algorithms

The computer vision project was the first Crowd Research deployment, and had the least structure: in the initial weeks, the aim of the project was intentionally kept vague and open to exploration. Participants spread out to find and summarize recent computer vision papers, then began an iterative process of formulating project proposals based on the review. Peers and the PI reviewed these proposals weekly. Eventually the PI aligned everyone on one team's proposal for integrating human workers with black-box classifiers to optimize performance at a certain crowdsourcing dollar cost.

With this new focus, participants developed datasets and evaluation procedures. However, many participants grew discouraged because their proposals were not selected. Some teams stepped up their work and became more collaborative, but others became far less active. This observation led the HCI and data science projects to keep project ideation more collective and less parallel, avoiding the abrupt cutoff of all but a single idea. In the final phase, the different groups worked in parallel to build interfaces, implement machine classifiers and perform experiments. The group published a work-in-progress poster at HCOMP 2015 with 54 authors (Veit et al. 2015).

5.1.2 HCI: The Daemo Crowdsourcing Platform

The HCI project, which created a new paid crowdsourcing platform, spent its initial weeks need finding by interviewing workers and requesters, then iteratively flared, focused, and rapidly prototyped research ideas. The crowd led ideation with feedback from the PI. As the research ideas solidified, participants self-selected which to participate in, each under different DRIs: the design and engineering of the platform—called Daemo—a new reputation system for Daemo, and an open governance structure for Daemo. These groups iterated on interaction design, engineering, user study design and analysis, and writing, again each under DRIs. After 12 weeks, the PI onboarded a second cohort of participants who joined the first group, continued work, and collectively published a work-in-progress poster at UIST 2015 with 70 authors (Gaikwad et al. 2015).

The group continued to work and submitted an integrated Daemo paper to CHI with 50 authors, but it was rejected principally for covering too many research thrusts in one paper. The group onboarded a third cohort a few months later and split the CHI submission into multiple papers. They published a full paper on the Boomerang reputation system at UIST 2016 with 37 authors (Gaikwad et al. 2016), and a full paper on Crowd Guilds at CSCW 2017 with 28 authors (Whiting et al. 2017). Daemo has launched in private beta (Gaikwad et al. 2017), and a paper based on data its workers collected won the best dataset paper award at EMNLP 2016 (Rajpurkar et al. 2016). The group continues work to launch Daemo publicly.

5.1.3 Data Science: Testing the Wisdom of Crowds at Scale

The data science project began with a literature review. Each participant found and summarized papers about the wisdom of crowds, extracting metadata about the task, the sample size, and the aggregation method. This produced 144 unique papers. Participants then synthesized 190 domains (e.g., calorie estimation, sports game prediction), which they narrowed by popular vote and PI input down to 50.

The PI next provided a question template that would allow the experiment to be deployed at scale. Participants curated 20 questions for each domain, including any images or audio clips, and committed them to a GitHub repository.

A small team of highly-motivated participants coded the experimental infrastructure to deploy these questions, and the group ran a pilot experiment. Participants analyzed the results from whatever angle seemed most interesting to them, and submitted a short report outlining their findings. The group published a work-in-progress poster at UIST 2015 with 60 authors (Mysore et al. 2015). The final paper is currently in preparation.

5.2 *The Crowd Led Projects' Ideation, Execution, and Writing*

Where did the research insights come from? We inductively generated themes using the milestone submissions each week from the HCI project, labeled each submission with a theme, and coded related themes across weeks.

The final research directions trace back to the crowd's early brainstorms (Fig. 7). These themes evolved into three ideas, two of which are now published. One set of themes in the first week, orange in Fig. 7 (low wages, uncertain payment, feelings of powerlessness among workers, and feelings of powerlessness among requesters), evolved into Crowd Guilds at CSCW (Whiting et al. 2017) and Daemo's open

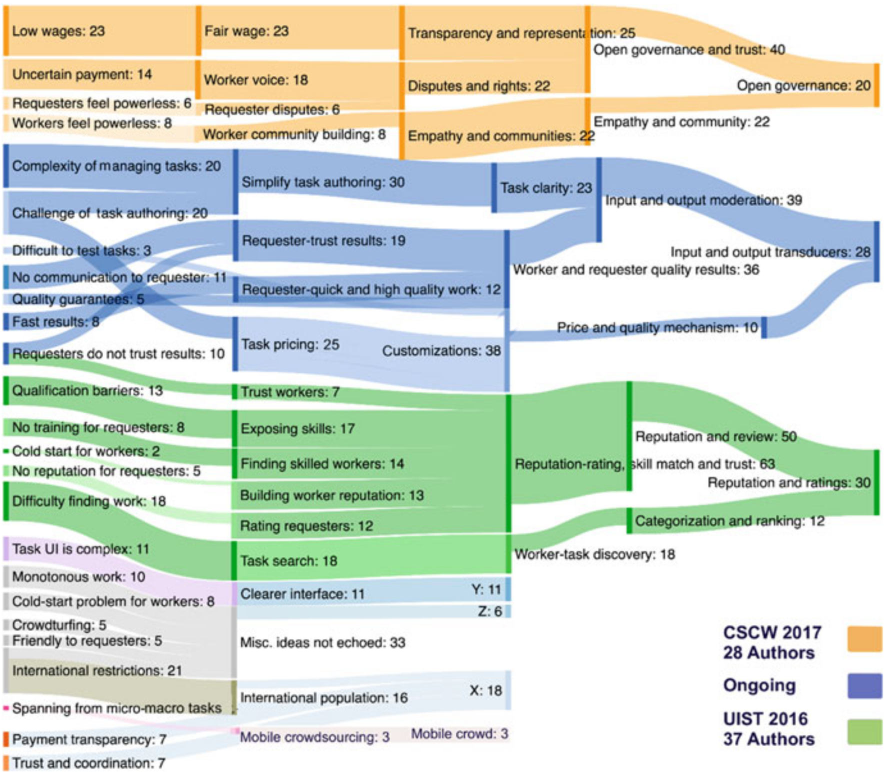


Fig. 7 The HCI crowd produced and iterated ideas over 7 weeks to develop three efforts, two of which are published papers now (orange and green). Darker shades were shared in the weekly meeting. Numbers report how many submissions proposed each idea

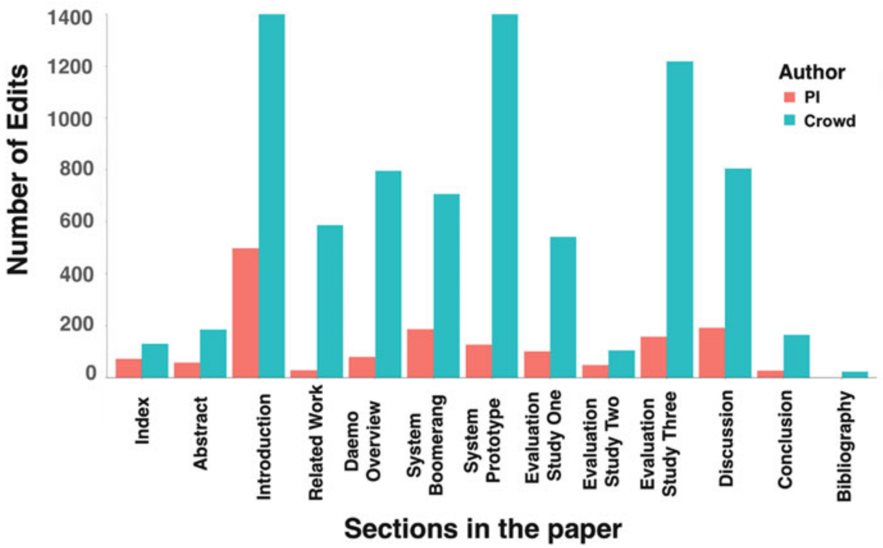


Fig. 8 The crowd led paper writing. In this submission, the crowd made 84% of paper edits, and the PI 16%

governance strategy. The green ideas in Fig. 7 included a lack of trust in result quality, no training for requesters, and qualification barriers. These evolved into a redesign of reputation systems, producing Boomerang at UIST (Gaikwad et al. 2016). The crowd made other suggestions that the PI chose not to echo back to the crowd, lighter-shaded in Fig. 7. Typically the PI did not echo ideas that were already played out in the research literature, for example the creation of a mobile crowdsourcing platform, or that did not constitute research goals, like addressing international restrictions for working on Mechanical Turk.

The crowd led paper writing as well (Fig. 8). We analyzed the edit history of the shared text editor used for one of the HCI project papers. Each edit represents insertion or deletion of a block of text. The crowd made 8360 edits (84%), while the principal investigator made 1580 edits (16%). The PI focused their edits mainly on the sections that frame the paper, such as the Introduction (Fig. 8). We compared this distribution to five papers in similar venues by the same PI but with their traditional Ph.D. students. On average, the Ph.D. students made 85% of the edits, and the PI 15% ($\sigma = 7\%$). So, this distribution is consistent with the PI's usual writing patterns.

Overall, Crowd Research enabled and empowered the crowd to choose domains of interest and lead diverse efforts. As one participant echoed via a survey: "I really enjoyed the freedom to collaborate and try out different tasks. I initially thought I would be on the coding side, but I found myself leading my group on open gov [ernance] and design initiatives from which I was able to successfully communicate and learn."

5.3 *Participants Remained Active for Months*

For Crowd Research to be effective, its participants must stay dedicated for a long period of time: research does not happen overnight. We measured active participation via Slack activity, because it indicates ongoing investment in the effort: team milestone submissions allow hiding behind a single active member, but Slack participation is tagged to each individual.

After 10 weeks, 15% of sign-ups and 29% of those who had participated in Slack were still active in the HCI project (Fig. 9). Participants were occasionally inactive due to exams and life events. Across projects, crowd members exchanged 500,000 Slack messages (1700 per week per project) and participated in 190,000 minutes of video meetings.

Several months after the projects launched, we surveyed active and inactive participants. Participants (N = 173) self reported a median 10 hours per week (mean 15 hours), which is substantial on top of other courses. The most common self-reported reasons for dropout were the inability to catch up after exams (53% agreed), the level of time commitment (35%), and losing friends or teammates (17%). This ranking of reasons was consistent across the three projects.

For their part, the PIs' time commitment depended on each PI's advising style. Some focused only on group meetings, while others helped read submissions. In

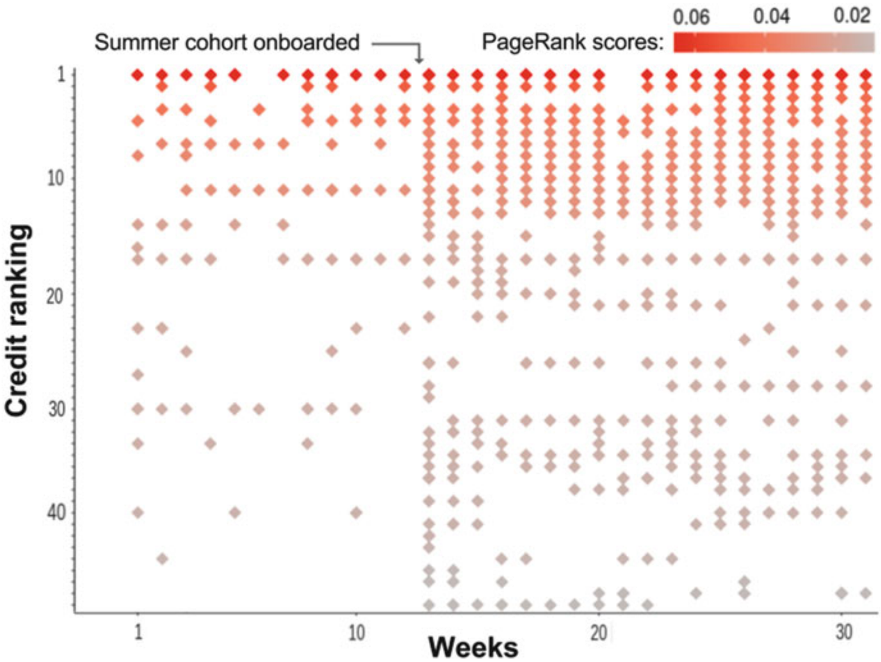


Fig. 9 Each row of dots represents a participant's weeks of active participation. The y-axis is the author order position

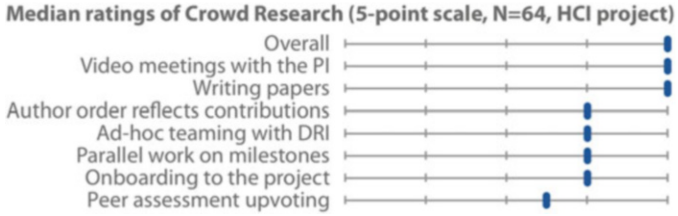


Fig. 10 Participants reported authentic research experiences, useful coordination strategies, and accurate credit

general, PIs spent 2 hours per week on the project: (1) a weekly 1-hour meeting with RAs to understand progress and design milestones; (2) a weekly 1-hour video advising meeting with the crowd; (3) sporadically helping with research-related questions over Slack. PIs engaged more heavily near deadlines.

Overall feedback was positive (Fig. 10). One participant shared: "This increases my interest in area of research. I learned many things from this project like writing research paper and skills like Angular JS and other frameworks, collaboration between team members and many more. It was a great enjoyable and educational learning experience."

5.4 *Crowd Research Provided New Routes for Access*

As evidence of access and upward mobility for traditional under-represented groups, we analyzed participants' self-reported age, gender, location, and affiliation from when they signed up. To understand whether participants had access to research experiences, we matched affiliations onto the Times Higher Education World University Rankings' subscale for research activity and influence (The Times Higher Education World University Rankings 2017), and locations at the country level onto a measure of GDP per capita (CIA 2016).

Most participants did not have prior access to research experiences. Seventy four percent were at institutions ranked below 500 worldwide. Sixty six percent were in countries ranked below 50 in GDP per capita.

Crowd research papers were substantially more diverse, in terms of authors' affiliation and current country, than others in the same top-tier venues (Fig. 11). We gathered all papers from CSCW 2017 and UIST 2016, where the papers were published, and compared the authors' affiliation rankings and country GDP per capita. The two Crowd Research papers had 57% and 58% of coauthors from universities ranked below 500 worldwide, vs. 12% and 11% of other papers in the venues (both $p < 0.001$). Likewise, the two Crowd Research papers had 42% and 35% of coauthors hailing from countries ranked below the top 50 in GDP per capita, vs. 2% and 6% for others papers in the two venues (both $p < 0.001$).

Participants leveraged PIs' recommendation letters to gain access to education and jobs. We surveyed all participants who received a letter from a PI, and

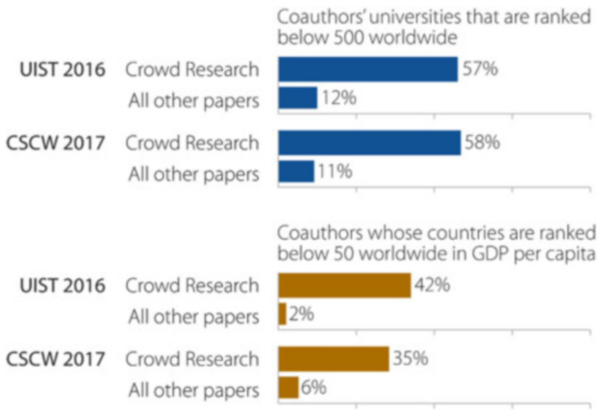


Fig. 11 Crowd Research paper authors were more diverse than others at the conferences they appeared at (all $p < 0.001$)

33 responded. Of these, 21 received an offer from an institution or a company that they applied to. These participants also sought other letters; however, a median of zero other letters (mean 0.37) were from universities or organizations that ranked above the top 500 worldwide. Thus, the Crowd Research PIs were the only recommenders from top-tier universities for many participants. These participants have since been admitted to undergraduate and graduate programs at schools including Stanford, UC Berkeley, Carnegie Mellon University, and MIT.

Feedback from participants emphasizes that they valued the access. As an undergraduate in India shared: "It provided me with the opportunity to associate myself with top research work, and these opportunities weren't available to us back home. It also allowed me to learn about the research methodology as practiced in universities such as Stanford and definitely went a long way in helping me secure admission."

Not everyone received admission—whether due to grades or insufficient contributions to Crowd Research to warrant a strong letter. They identified other benefits: "While involvement with Crowd Research has given momentum to my pursuit of the future of problem solving and work, its effect on my current career as a librarian has been uncertain. I am not troubled by this though, because I believe any short term opportunity costs will be made up by long term benefits of having the foundation laid by my Crowd Research experience. [. . .] Any concern I have for my own career is far outweighed by my interest in shaping the nature of individual contributions to society at a large scale. That said, I have been able to secure funding for several conferences, am in the process of writing a white paper."

5.5 Decentralized Credit Amplified Concrete Contributions

Participants felt that the author orders reflected their contributions (Median Likert 4/5, Fig. 10). Figure 12 plots the cumulative distribution of PageRank scores for the CHI 2015 submission. There is a clustering of low scores representing about 60% of

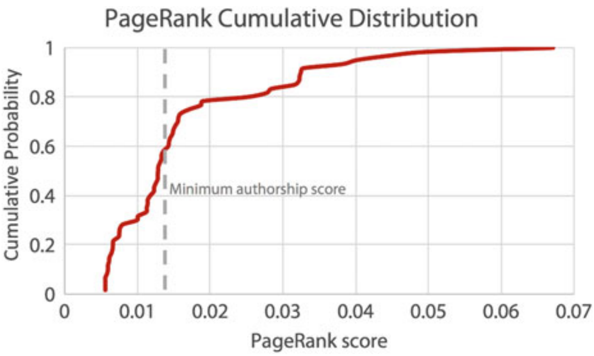


Fig. 12 A CDF of the PageRank scores for the paper. Only those active up through the paper submission were eligible (after about 1 year). Of those, 36 were included as coauthors, and 41 were below the threshold and included in Acknowledgments instead

Table 1 Regressions comparing the effects of participation behaviors on credit

Participation measure	Page rank score β_{PR}	Raw score β_{RS}	Difference $\beta_{PR} - \beta_{RS}$
Meetings present	0.069***	0.044*	0.026
Files uploaded	0.035**	0.029*	0.006
GitHub commits	0.017	−0.024*	0.041***
Slack messages	0.035*	0.112***	−0.077***
Self-organized meetings	0.024*	0.012	0.012
Was DRI (binary)	0.036***	0.006	0.030**
Weeks active	0.025*	0.014	0.011

Right column: compared to raw votes, PageRank increased the value of GitHub commits and DRI-ing, and decreased the value of talking on Slack. The median raw score was 0.06 [IQR (0.03, 0.15)], and the median distance between two adjacent authors was 0.003. The median PageRank score was 0.11 [IQR (0.03, 0.2)], and the median distance between two adjacent authors was 0.004. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

still-active accounts below the authorship threshold, and the remaining 40% more spread out amongst the higher scores.

But what effect did the networked credit allocation have on the author order? To investigate, we compared the initial raw credit scores and final PageRank-adjusted scores against logs of contributor behavior. We normalized the raw credit scores and the PageRank-adjusted scores to sum to 1.0. We then performed two multiple regressions, one predicting normalized raw score and one predicting PageRank score. Independent variables were observable participation behaviors, including Slack, GitHub, and weekly meetings, all standardized into z-scores. The regression coefficients β explain which behaviors were significantly correlated with changes in credit score. The right column of Table 1 highlights which features were significantly different between the raw credit and PageRank.

Relative to the raw score, PageRank lessened the impact of sending chat messages on Slack, and increased the impact of DRI-ing and committing code to GitHub (Table 1). This means that PageRank credit increased the effects of concrete contributions. These effects materially changed the author order. For example, one large team, who rarely interacted with the rest of the crowd, assigned nearly all of their credits to their team lead (Fig. 6 top). Raw votes placed the team's lead second in the overall author order. However, PageRank softened the effect of this link ring (Fig. 6 bottom), because others did not assign nearly as much credit to the team.

As one participant shared: "It's obvious that if you have such a credit system, someone would try to cheat his/her reputation, but I think our credit system worked very well." One point of frustration was last-minute contributors: "Some people just appear 2 days before the paper submission deadline and take over the Slack channel and talk a lot. Then the result is that they are up-voted and got into the author list." But most feedback was favorable: "The idea of peer-evaluation and the use of PageRank resulted in mostly fair and accurate results."

6 Discussion

The Deployment section focused on successes. Here, we reflect on the challenges of Crowd Research. Challenges are instructive: they teach us the limitations of the technique, unpredicted outcomes, and opportunities for future research.

6.1 How to Run a Bad Crowd Research Project

What lessons can be drawn for HCI and social computing? It can be more enlightening to discuss failure modes than successes. We offer a David Patterson-style list (Patterson 1994) of ways run a bad Crowd Research project:

- *Assume 100% followthrough. Participants are extremely motivated, and work on Crowd Research to the exclusion of everything else*. Even motivated contributors have jobs, exams, and lives. Milestones need to either utilize redundancy, or enforce deadlines and allow tasks to be reassigned if participants do not meet them.
- *Encourage competition. Let the best contributions rise to the top*. This defaults to a critical culture, leading to dropout. It is critical to establish norms for a positive, inclusive culture (Kiesler et al. 2012). In our case, switching from doubleblind feedback to an upvoting system, plus consistent PI communication, helped changed the norm.
- *Treat the crowd like incompetent undergraduates or mature graduate students*. Rote work leads to a lack of interest, but being too open-ended leaves many people behind. Balance the two through focused short-term milestones that encourage creativity.
- *Pick projects you would do with your current lab*. This underplays the benefits of the crowd. It is better to leverage scale and diversity to achieve more ambitious goals.

- *Assume that nobody will come into conflict*. Running a Crowd Research project feels like being in charge of a team or organization, giving rise to lots of progress but also interpersonal issues. This comes to a head especially around credit.

6.2 Limitations

One common question about Crowd Research is whether the effort is worth the PI's time investment of 2–3 hours per week. This is certainly higher than a single once-a-week meeting. However, we would tend not to make a direct effort comparison. First, the three crowd research projects were more ambitious than typical projects in their respective labs, for example building a new crowdsourcing platform or running hundreds of experiments, making them difficult to compare to traditional papers. This was by design: we sought projects that capitalized on having a crowd. Second, the stated goal of Crowd Research is enabling access, not publishing more papers per hour. Empowering the crowd seemed worth an incremental extra time commitment.

A second critique is how much success can be attributed to the PI rather than the crowd. There was certainly PI-driven variation across projects: with Computer Vision the PI went on sabbatical and the project stopped after a WIP; with HCI the project was sustained through publication. However, the ideas themselves were crowd-driven (Fig. 7), and like with advising traditional graduate students, most often the PI was helping filter bad ideas and amplify good ideas rather than propose all the ideas themselves.

A third question is to what extent any prestige associated with the universities were responsible for success. It seems likely that these names increased initial enrollment. However, the most popular project, Computer Vision, was the only one without a PI from a Top 15 university. This suggests that interest area rather than university name may have a substantial effect.

6.3 Disagreements and Biases

Like any distributed team, conflicts broke out (Hinds and Bailey 2003; Hinds and Mortensen 2005). Most commonly, these issues arose between participants: objections over the influence that someone was wielding, misaligned values (e.g., "talkers" vs. "doers"), disagreements on research decisions, or second-guessing of intentions. The PI or RAs diffused these situations, but they took an emotional toll on everyone involved. While most participants felt that author ordering was helpful, it was also a source of tension because votes were kept private. No matter how high someone was on the author list, we would hear complaints that they should have been ranked higher, or others ranked lower. In rare cases, participants publicly called each other out, which sowed tension and negatively affected trust.

A global project must also contend with cultural differences. The upside of cultural diversity is increased creativity and satisfaction (Kulkarni et al. 2015; Stahl et al. 2010). The downside is that diversity can lead to ethnocentrism, implicit and explicit biases (Cramton and Hinds 2004). Different cultures idealize different behaviors (Li 2003; Tsai 2007). Different cultures may also exhibit biases in how they treat women or other groups. If not carefully managed, cultural differences may drive out qualified participants or undervalue their contributions. It is nearly impossible to remove implicit biases from Crowd Research participants' credit evaluations of each other. Future work will measure the extent of these biases and identify ways to counteract them.

6.4 Future Work

In the future, we hope to expand Crowd Research beyond Computer Science topics. In addition, we will make its suite of tools more turnkey so that any interested group can easily spin up a project. Finally, we hope to perform a longitudinal analysis or randomized trial to directly examine the long-term effects of participation.

7 Conclusion

This chapter presents Crowd Research and an analysis of its 2 year long deployment. Crowd Research introduces a crowdsourcing technique for coordinating a large group of people in an open-ended research exploration, and a system for decentralized credit distribution. It enabled access to over 1500 people worldwide to collaborate online in the pursuit of open-ended research. Utilizing Crowd Research, participants have built real-world systems, co-authored papers for top-tier conferences and have gone on to further careers in research.

Crowd Research represents a new form of knowledge production—one that leverages the diversity and scale of the internet to pursue projects that might be challenging in traditional laboratory environments. We believe that if Crowd Research and similar techniques successfully enable global access to training and mentorship experiences, they will help grow a new generation of scientists, humanists, and engineers that increase diversity in the scientific workforce. We envision that this generation could work collectively to resolve some of the biggest unanswered questions of our time.

Acknowledgements We thank over 1500 members of the Stanford Crowd Research Collective community for their contributions. This work was supported by Office of Naval Research awards N00014-16-1-2894 and N00014-15-1-2711, Institute for Scalable Scientific Data Management at UCSC and Los Alamos National Laboratory, Toyota, and the Hasso-Plattner Design Thinking Research Program.

References

Baeza-Yates, R. A., Castillo, C., López, V., Shubik, M., Hopcroft, J., & Sheldon, D. (2007). Pagerank increase under different collusion topologies. In *AIRWeb* (Vol. 5, pp. 68–81). New York: Springer.

Bianchini, J. A. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads*. Washington: National Academies.

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

Bowen, W. G., & Bok, D. (2016). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton: Princeton University Press.

Budhathoki, N. (2016). *Who are the mappers and why do they map in OpenStreetMap*. https://www.youtube.com/watch?v=LvakiUOsDrM

Chen, Y., Oney, S., & Lasecki, W. S. (2016). Towards providing on-demand expert support for software developers. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 3192–3203). New York: ACM.

CIA. (2016). *The world factbook*. https://www.cia.gov/library/publications/the-world-factbook/

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popovic, Z., & Players, F. (2010). Predicting protein structures with a multiplayer online game. *Nature, 466*(7307), 756–760.

Cox, J., Eun Young, O., Simmons, B., Lintott, C., Masters, K., Greenhill, A., Graham, G., & Holmes, K. (2015). Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Computing in Science & Engineering, 17*(4), 28–41.

Cramton, C. D., & Hinds, P. J. (2004). Subgroup dynamics in internationally distributed teams: Ethnocentrism or cross-national learning? *Research in Organizational Behavior, 26*, 231–263.

Cranshaw, J., & Kittur, A. (2011). The polymath project: Lessons from a successful online collaboration in mathematics. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1865–1874). New York: ACM.

Cronje, R., Rohlinger, S., Crall, A., & Newman, G. (2011). Does participation in citizen science improve scientific literacy? A study to compare assessment methods. *Applied Environmental Education & Communication, 10*(3), 135–145.

Dillahunt, T. R., Chen, B., & Teasley, S. (2014). Model thinking: Demographics and performance of MOOC students unable to afford a formal education. In *Proceedings of the first ACM conference on learning@ scale conference* (pp. 145–146). New York: ACM.

Dillahunt, T. R., Ng, S., Fiesta, M., & Wang, Z. (2016). Do massive open online course platforms support employability? In *Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing* (pp. 232–243). New York: ACM.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology, 60*(11), 2229–2243.

Dow, S., Kulkarni, A., Klemmer, S., & Hartmann, B. (2012). Shepherding the crowd yields better work. In *Proceedings of the ACM 2012 conference on computer supported cooperative work* (p. 1013). New York: ACM.

Eric Landrum, R., & Nelsen, L. R. (2002). The undergraduate research assistantship: An analysis of the benefits. *Teaching of Psychology, 29*(1), 15–19.

Gaikwad, S. 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. S., Sehgal, V., Lin, J., Kristianto, L., Regino, J., Chhibber, N., Majeti, D., Sharma, S., Mananova, K., Dhakal, D., Dai, W., Purynova, V., Sandeep, S., Chandrakanthan, V., Sarma, T., Matin, S., Nassar, A., Nistala, R., Stolzoff, A., Milland, K., Mathur, V., Vaish, R., & Bernstein, M. S. (2016). Boomerang: Rebounding the consequences of reputation feedback on crowdsourcing platforms. In *Proceedings of the 29th annual ACM symposium on user interface software and technology (UIST '16)*. New York: ACM.

Gaikwad, S. S., Morina, D., Nistala, R., Agarwal, M., Cossette, A., Bhanu, R., Savage, S., Narwal, V., Rajpal, K., Regino, J., Mithal, A., Ginzberg, A., Nath, A., Ziulkoski, K. R., Cossette, T., Gamage, D., Richmond-Fuller, A., Suzuki, R., Herrejón, J., Le, K., Flores-Saviaga, C., Thilakarathne, H., Gupta, K., Dai, W., Sastry, A., Goyal, S., Rajapakshe, T., Abolhassani, N., Xie, A., Reyes, A., Ingle, S., Jaramillo, V., Godinez, M., Angel, W., Toxtli, C., Flores, J., Gupta, A., Sethia, V., Padilla, D., Milland, K., Setyadi, K., Wajirasena, N., Batagoda, M., Cruz, R., Damon, J., Nekkanti, D., Sarma, T., Saleh, M., Gongora-Svartzman, G., Bateni, S., Barrera, G. T., Pena, A., Compton, R., Aariff, D., Palacios, L., Ritter, M. P., Nisha, K. K., Kay, A., Uhrmeister, J., Nistala, S., Esfahani, M., Bakiu, E., Diemert, C., Matsumoto, L., Singh, M., Patel, K., Krishna, R., Kovacs, G., Vaish, R., & Bernstein, M. (2015). Daemo: A self-governed crowdsourcing marketplace. In *Adjunct proceedings of the 28th annual ACM symposium on user interface software & technology* (pp. 101–102). New York: ACM.

Gaikwad, S. S., Whiting, M. E., Gamage, D., Mullings, C. A., Majeti, D., Goyal, S., Gilbee, A., Chhibber, N., Ginzberg, A., Richmond-Fuller, A., Matin, S., Sehgal, V., Sarma, T. S., Nasser, A., Ballav, A., Regino, J., Zhou, S., Mananova, K., Srinivas, P., Ziulkoski, K., Dhakal, D., Stolzoff, A., Niranga, S. S., Salih, M. H., Sinha, A., Vaish, R., & Bernstein, M. S. (2017). The daemo crowdsourcing marketplace. In *Companion of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 1–4). New York: ACM.

Gerstner, C. R., & Day, D. V. (1997). Meta-analytic review of leader-member exchange theory: Correlates and construct issues. *Journal of Applied Psychology, 82*(6), 827–844.

Gilbert, E. (2013). Widespread underprovision on Reddit. In *Proceedings of the 2013 conference on computer supported cooperative work* (pp. 803–808). New York: ACM.

Gowers, T., & Nielsen, M. (2009). Massively collaborative mathematics. *Nature, 461*(7266), 879–881.

Gurin, P., Dey, E., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review, 72*(3), 330–367.

Haizi, Y., Deka, B., Talton, J. O., & Kumar, R. (2016). Accounting for taste: Ranking curators and content in social networks. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2383–2389). New York: ACM.

Hathaway, R. S., Nagda, B. A., & Gregerman, S. R. (2002). The relationship of undergraduate research participation to graduate and professional education pursuit: An empirical study. *Journal of College Student Development, 43*(5), 614.

Hinds, P. J., & Bailey, D. E. (2003). Out of sight, out of sync: Understanding conflict in distributed teams. *Organization Science, 14*(6), 615–632.

Hinds, P. J., & Mortensen, M. (2005). Understanding conflict in geographically distributed teams: The moderating effects of shared identity, shared context, and spontaneous communication. *Organization Science, 16*(3), 290–307.

Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences, 101*(46), 16385–16389.

Hunter, A.-B., Laursen, S. L., & Seymour, E. (2007). Becoming a scientist: The role of undergraduate research in students' cognitive, personal, and professional development. *Science Education, 91*(1), 36–74.

Irani, L. C., & Silberman, M. (2013). Turkopticon: Interrupting worker invisibility in amazon mechanical turk. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 611–620). New York: ACM.

Jordan Raddick, M., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., Szalay, A. S., & Vandenberg, J. (2009). *Galaxy zoo: Exploring the motivations of citizen science volunteers*. arXiv preprint arXiv:0909.2925.

Kevin Eagan, M., Hurtado, S., Chang, M. J., Garcia, G. A., Herrera, F. A., Garibay, J. C., & Garibay, J. C. (2013). Making a difference in science education: The impact of undergraduate research programs. *American Educational Research Journal, 50*(4), 683–713.

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popovic, Z., Jaskolski, M., & Baker, D. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology, 18*(10), 1175–1177.

Kiesler, S., Kraut, R., Resnick, P., & Kittur, A. (2012). Regulating behavior in online communities. In *Building successful online communities: Evidence-based social design*. Cambridge: MIT.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. (2013). The future of crowd work. In *Proceedings of the 2013 conference on computer supported cooperative work* (p. 1301). New York: ACM.

Kittur, A., Smus, B., Khamkar, S., & Kraut, R. E. (2011). Crowdforge: Crowdsourcing complex work. In *Proceedings of the 24th annual ACM symposium on user interface software and technology* (pp. 43–52). New York: ACM.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. In *Proceedings of the second ACM conference on learning@ scale* (pp. 57–66). New York: ACM.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM), 46*(5), 604–632.

Kulkarni, C. E., Bernstein, M. S., & Klemmer, S. R. (2015). PeerStudio: Rapid peer feedback emphasizes revision and improves performance. In *Proceedings of the second (2015) ACM conference on learning@ scale* (pp. 75–84). New York: ACM.

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., Cheng, J., Koller, D., & Klemmer, S. R. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction (TOCHI), 20*(6), 33.

Lakhani, K. R., Lifshitz-Assaf, H., & Tushman, M. (2013). Open innovation and organizational boundaries: Task decomposition, knowledge distribution and the locus of innovation. In *Handbook of economic organization: Integrating economic and organizational theory* (pp. 355–382). Cheltenham: Edward Elgar.

Land, K., Slosar, A., Lintott, C., Andreescu, D., Bamford, S., Murray, P., Robert, N., Jordan Raddick, M., Schawinski, K., Szalay, A., Thomas, D., & Vandenberg, J. (2008). Galaxy zoo: The large-scale spin statistics of spiral galaxies in the sloan digital sky survey. *Monthly Notices of the Royal Astronomical Society, 388*(4), 1686–1692.

Lashinsky, A. (2012). *Inside apple: How America's most admired–and secretive–company really works*. London: John Murray.

LaToza, T. D., Ben Towne, W., Adriano, C. M., & Van Der Hoek, A. (2014). Microtask programming: Building software with a crowd. In *Proceedings of the 27th annual ACM symposium on user interface software and technology* (pp. 43–54). New York: ACM.

Law, E., Wiggins, A., Gray, M. L., & Williams, A. (2017). Crowdsourcing as a tool for research: Implications of uncertainty. In *Proceedings of the 20th ACM conference on computer-supported cooperative work & social computing (CSCW '17)*. New York: ACM.

Lee, J., Kladwang, W., Lee, M., Cantu, D., Azizyan, M., Kim, H., Limpaecher, A., Gaikwad, S., Yoon, S., Treuille, A., et al. (2014). RNA design rules from a massive open laboratory. *Proceedings of the National Academy of Sciences, 111*(6), 2122–2127.

Li, J. (2003). US and Chinese cultural beliefs about learning. *Journal of Educational Psychology, 95*(2), 258.

Lixiu, Y., Kittur, A., & Kraut, R. E. (2014). Distributed analogical idea generation: Inventing with crowds. In *Proceedings of the 32nd annual ACM conference on human factors in computing systems* (pp. 1245–1254). New York: ACM.

Lopatto, D. (2004). Survey of undergraduate research experiences (SURE): First findings. *Cell Biology Education, 3*(4), 270–277.

Mao, A., Mason, W., Suri, S., Watts, D. J., & Malone, T. W. (2016). An experimental study of team size and performance on a complex task. *PLoS One, 11*(4), e0153048.

Martin, D., Hanrahan, B. V., O'Neill, J., & Gupta, N. (2014). Being a turker. In *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing* (pp. 224–235). New York: ACM.

Miller, R. C., Zhang, H., Gilbert, E., & Gerber, E. (2014). Pair research: Matching people for collaboration, learning, and productivity. In *Proceedings of the 17th ACM conference on computer supported cooperative work & social computing* (pp. 1043–1048). New York: ACM.

Mysore, A. S., Yaligar, V. S., Ibarra, I. A., Simoiu, C., Goel, S., Arvind, R., Sumanth, C., Srikantan, A., Bhargav, H. S., Pahadia, M., Dobha, T., Ahmed, A., Shankar, M., Agarwal, H., Agarwal, R., Anirudh-Kondaveeti, S., Arun-Gokhale, S., Attri, A., Chandra, A., Chilukur, Y., Dharmaji, S., Garg, D., Gupta, N., Gupta, P., Jacob, G. M., Jain, S., Joshi, S., Khajuria, T., Khillan, S., Konam, S., Kumar-Kolla, P., Loomba, S., Madan, R., Maharaja, A., Mathur, V., Munshi, B., Nawazish, M., Neehar-Kurukunda, V., Nirmal-Gavarraju, V., Parashar, S., Parikh, H., Paritala, A., Patil, A., Phatak, R., Pradhan, M., Ravichander, A., Sangeeth, K., Sankaranarayanan, S., Sehgal, V., Sheshan, A., Shibiraj, S., Singh, A., Singh, A., Sinha, P., Soni, P., Thomas, B., Varma-Dattada, K., Venkataraman, S., Verma, P., & Yelurwar, I. (2015). Investigating the "wisdom of crowds" at scale. In *Adjunct proceedings of the 28th annual ACM symposium on user interface software & technology (UIST '15 adjunct)* (pp. 75–76). New York: ACM.

Nebeling, M., To, A., Guo, A., de Freitas, A. A., Teevan, J., Dow, S. P., & Bigham, J. P. (2016). WearWrite: Crowd-assisted writing from smartwatches. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 3834–3846). New York: ACM.

Oliveira, N., Jun, E., & Reinecke, K. (2017). Citizen science opportunities in volunteer-based online experiments. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6800–6812). New York: ACM.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, 6251.

Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton: Princeton University Press.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the web*. Stanford: Stanford InfoLab.

Pandey, V., Amir, A., Debelius, J., Hyde, E. R., Kosciolek, T., Knight, R., & Klemmer, S. (2017). Gut instinct: Creating scientific theories with online learners. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 6825–6836). New York: ACM.

Patterson, D. A. (1994). How to have a bad career in research/academia. In *Keynote, 1994 USENIX symposium on operating system design and implementation*.

Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 conference on empirical methods in natural language processing*.

Reinecke, K., & Gajos, K. Z. (2015). LabintheWild: Conducting large-scale online experiments with uncompensated samples. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing* (pp. 1364–1378). New York: ACM.

Retelny, D., Robaszkiewicz, S., To, A., Lasecki, W. S., Patel, J., Rahmati, N., Doshi, T., Valentine, M., & Bernstein, M. S. (2014). Expert crowdsourcing with flash teams. In *Proceedings of the 27th annual ACM symposium on user interface software and technology* (pp. 75–85). New York: ACM.

Russakovsky, O., Li, L.-J., & Li, F.-F. (2015). Best of both worlds: Human-machine collaboration for object annotation. In *IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 2121–2131). Boston: IEEE.

Russell, S. H., Hancock, M. P., & McCullough, J. (2007). Benefits of undergraduate research experiences. *Science, 316*(5824), 548–549.

Sarsons, H. (2015). *Gender differences in recognition for group work*. Working Paper, Harvard University.

Shen, H.-W., & Barabási, A.-L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences, 111*(34), 12325–12330.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution, 24*(9), 467–471.

Stahl, G. K., Maznevski, M. L., Voigt, A., & Jonsen, K. (2010). Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of International Business Studies, 41*(4), 690–709.

Stevens, M., Vitos, M., Altenbuchner, J., Conquest, G., Lewis, J., & Haklay, M. (2014). Taking participatory citizen science to extremes. *IEEE Pervasive Computing, 13*(2), 20–29.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation, 142* (10), 2282–2292.

Sun, Y., Yintao, Y., & Han, J. (2009). Ranking-based clustering of heterogeneous information networks with star network schema. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 797–806). New York: ACM.

Suzuki, R., Salehi, N., Lam, M. S., Marroquin, J. C., & Bernstein, M. S. (2016). Atelier: Repurposing expert crowdsourcing tasks as micro-internships. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2645–2656). New York: ACM.

Teevan, J., Iqbal, S. T., & von Veh, C. (2016). Supporting collaborative writing with microtasks. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 2657–2668). New York: ACM.

The Times Higher Education World University Rankings. (2017). https://www.timeshighereducation.com/world-university-rankings

Toegel, G., & Conger, J. A. (2003). 360-degree assessment: Time for reinvention. *Academy of Management Learning & Education, 2*(3), 297–311.

Tomlinson, B., Ross, J., Andre, P., Baumer, E., Patterson, D., Corneli, J., Mahaux, M., Nobarany, S., Lazzari, M., Penzenstadler, B., et al. (2012). Massively distributed authorship of academic papers. In *Extended abstracts of the 2012 ACM annual conference on human factors in computing systems* (pp. 11–20). New York: ACM.

Tsai, J. L. (2007). Ideal affect: Cultural causes and behavioral consequences. *Perspectives on Psychological Science, 2*(3), 242–259.

Valentine, M. A., Retelny, D., To, A., Rahmati, N., Doshi, T., & Bernstein, M. S. (2017). Flash organizations: Crowdsourcing complex work by structuring crowds as organizations. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 3523–3537). New York: ACM.

Veit, A., Wilber, M. J., Vaish, R., Belongie, S. J., Davis, J., Anand, V., Aviral, A., Chakrabarty, P., Chandak, Y., Chaturvedi, S., Devaraj, C., Dhall, A., Dwivedi, U., Gupte, S., Sridhar, S. N., Paga, K., Pahuja, A., Raisinghani, A., Sharma, A., Sharma, S., Sinha, D., Thakkar, N., Bala Vignesh, K., Verma, U., Abhishek, K., Agrawal, A., Aishwarya, A., Bhattacharjee, A., Dhanasekar, S., Gullapalli, V. K., Gupta, S., Chandana, G., Jain, K., Kapur, S., Kasula, M., Kumar, S., Kundaliya, P., Mathur, U., Mishra, A., Mudgal, A., Nadimpalli, A., Nihit, M. S., Periwal, A., Sagar, A., Shah, A., Sharma, V., Sharma, Y., Siddiqui, F., Singh, V., Abhinav, S., Tambwekar, P., Taskin, R., Tripathi, A., & Yadav, A. D. (2015). On optimizing human-machine task assignments. HCOMP 2015 extended abstracts.

Whiting, M. E., Gamage, D., Snehalkumar (Neil) S Gaikwad, Gilbee, A., Goyal, S., Ballav, A., Majeti, D., Chhibber, N., Richmond-Fuller, A., Vargus, F., Sarma, T. S., Chandrakanthan, V., Moura, T., Salih, M. H., Kalejaiye, G. B. T., Ginzberg, A., Mullings, C. A., Dayan, Y., Milland, K., Orefice, H., Regino, J., Parsi, S., Mainali, K., Sehgal, V., Matin, S., Sinha, A., Vaish, R., & Bernstein, M. S. (2017). Crowd guilds: Worker-led reputation and feedback on crowdsourcing platforms. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing (CSCW '17)* (pp. 1902–1913). New York: ACM.

Zhang, H., Easterday, M. W., Gerber, E. M., Lewis, D. R., & Maliakal, L. (2017). Agile research studios. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing – CSCW '17* (pp. 220–232). New York: ACM.

Zydney, A. L., Bennett, J. S., Shahid, A., & Bauer, K. W. (2002). Impact of undergraduate research experience in engineering. *Journal of Engineering Education, 91*, 151–157.

Redesigning Social Organization for Accelerated Innovation in the New Digital Economy: A Design Thinking Perspective

Ade Mabogunje, Neeraj Sonalkar, and Larry Leifer

Abstract We now appear to be in the full grip of the media transformation from paper-based media to a digital-based media. This evolution in mobility of information (experiences) has occurred alongside the mobility of matter and labor (goods and services, mass and heat), all of which have come about as a result of evolution in technologies of encryption, computation, communication, representation, sensing, and transportation. All these changes have contributed to a market environment that is more open, connected, complex, and dynamic, and to corporate and civic organizational configurations that are overwhelmed and slow to adapt to these changes. In the Hasso Plattner Design Thinking Research program, we have been observing these changes, and developing solutions to accelerate the rate of innovation in the new digital economy. Our work has led us to focus on the design team, the design coach, and the instrumented design space as the new unit of knowledge work, as opposed to the individual employees and line manager. This new unit is larger than the individual and so can take in more information. It is smaller than the typical organizational group or department, so it is faster to act and more agile. And the data rich and computational nature of the instrumented space, means that the technology can be considered a bona fide member of the design team. The variety of organizational structures now possible as well as the way these structures need to change in very short time frames has made it necessary to develop a biological metaphor of the organization as an organism that can fold, unfold, and refold as it adapts rapidly to a fast-changing environment. This radical shift from the hierarchical, clockwork, command and control organizations of the industrial age, will be explored with a view to showing alternative redesign of social organizations and the means to accomplish the requisite sociological, psychological, and technological transformations effectively.

A. Mabogunje (✉) · N. Sonalkar · L. Leifer
Center for Design Research, Stanford University, Stanford, CA, USA
e-mail: ade@stanford.edu; sonalkar@stanford.edu; leifer@cdr.stanford.edu

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_8

1 Introduction

Organizations are increasingly embracing design thinking methodology as a way of structuring their innovation endeavors. Leading business journals such as Harvard Business Review proclaim design thinking as a key methodology for organizational innovation (Kolko 2015). There have been a number of case studies documenting groups within organizations practicing design thinking with varying levels of success (Köppen et al. 2016). A deeper study of these case studies reveals that design thinking is used in three ways—as a workshop method to introduce employees to new ways of creative thinking, as a product development methodology that is user-centric, or as a methodology to drive change in organizational productivity and responsiveness. These case studies also reveal a number of challenges in implementing design thinking within an organization such as strategic directives from top management clashing with user-centric insights by product teams, organizational silos circumscribing design thinking and converting it into its own silo, and individuals who develop a design mindset feeling restricted by existing organizational routines (Schmiedgen et al. 2016). We too have faced difficulties implementing high performance design teams in industry, not because companies do not seek design team performance, but rather because the broader organizational context in which these teams operate does not fully support high performing design thinking teams. Thus, there is an imperative to understand the role of social organization and its influence on design thinking and ask the question—what kind of an organization is required to realize the full potential of high performance design teams?

In this book chapter, we will begin by outlining the critical relationship between a design team's context and the team's performance. Next, we will describe in greater detail the nature of current social organizations in which design teams are embedded and examine their advantages and disadvantages. Following this, we will propose a new form of social organization centered around instrumented design spaces and consisting of teams that can form and reform to act and respond faster to market needs and technology opportunities. Next, we will discuss the merits of this new form, followed by an example. We will conclude the chapter by examining the range of changes that people need to make to transform existing social organizations into this new form of a social organization.

2 Design Teams Are Context-Dependent

It important to consider the nature of the social organization in which a design team is embedded for four main reasons. First, a design team's performance occurs in the broader field of a market environment. Second, the nature of the social organization determines access to resources that a team needs in order to perform design work. Third, the nature of social organization dictates an assumption of roles which define what behavior is acceptable to the organization irrespective of its effectiveness in the market. Fourth, the nature of the social organization determines the identity or group to which an individual feels he or she belongs.

A design team's performance occurs in the broader field of market environment. In the digital economy, the market is an ever-changing environment. New companies are coming into this environment and new devices are being invented in this environment. Current customers that are using a product, are constantly adapting it to their purpose so that a team has to observe them through need-finding to identify an opportunity for growth. If a company does not act on such opportunities, its competitors might, and in so doing disrupt the market share of that company. High-performance design team needs to operate in such rapidly changing market conditions, discover opportunities, and create products that address user needs and drive company's business growth. The social organization within a company mediates a team's interactions with the market. This mediation occurs in several ways, which leads to the next point.

The nature of social organization determines access to resources that a team needs to perform design work. In order to be a high-performance design team, the members need access to resources such as capital, information and additional talent. These can be constrained by the nature of social organization in such a way that it hinders high performance design behaviors. For example, access to information can be restricted by middle management through a permission process that seeks to safeguard role legitimacy rather than promote creative teamwork.

The nature of social organization dictates an assumption of roles which define what behavior is acceptable to the organization irrespective of its effectiveness in the market. Once a team is a part of an organization, each team member develops an assumption of the role that he or she is actually playing out (Katz and Kahn 1978). The social organization reinforces this role through mimesis, narratives, and incentives. This role then becomes a key driver of individual and team behavior rather than the market.

Finally, the nature of social organization determines the occupational, professional, or class (owner, manager, employee, contractor) identity that an individual feels he or she belongs to. This identity further influences team interactions and either constrains or facilitates collaborative relationships with others in the market. One of the researchers who first drew attention to this phenomenon was Anna Lee Saxenian in her book, "Regional Advantage," where she compared the Silicon Valley to companies on Route 128 on the East Coast of the US. Her study showed that California's Silicon Valley was able to keep up with the fast pace of technological progress during the 1980s, while the vertically integrated firms of the Route 128 beltway fell behind (Saxenian 1996). She argued that the key was Silicon Valley's decentralized organizational form, non-proprietary standards, and tradition of cooperative exchange (sharing information and outsourcing for component parts), in opposition to hierarchical and independent industrial systems in the East Coast. This led to Silicon Valley engineers identifying with the entire regional ecosystem rather than one company, which in turn accelerated information sharing and worker mobility, while the Route 128 engineers identified themselves with only their company and were limited in their interactions to people within that company.

Thus, in order to foster information and idea sharing type team interactions, we need to consider the nature of social organizations in which such teams are expected to operate, and if necessary, we may need to redesign them.

3 The Hierarchical Organization

Organizations can be considered a form of ordering planned action. Human beings organize themselves so that a particular action occurs to achieve the desired goal more efficiently. The dominant form through which humans organize is that of a hierarchy with predetermined lines of communications. Another name for this form of organizing is command and control (Fig. 1).

The "command and control" structure comes out of military history where we as humans organized ourselves for campaigns. It typically consists of people organized into different strata, with power distributed between these strata. There is a stratum where people hold more power than the layer below it, and that layer holds more power than layer below it, and so on. This way, power distribution comes in the picture when we talk about this form of organization.

The "command and control" hierarchy could also be considered a "thinking and doing" hierarchy, or "planning and execution" hierarchy, where there is a separation between thinking and doing. Take for example, in 1969, Drucker wrote:

> . . . For business, during the last thirty years, has had to face, on a much smaller scale, the problem government now faces: the incompatibility between "governing" and "doing." Business management learned that the two have to be separated, and that the top organ, the decision maker, has to be detached from "doing." Otherwise he does not make decisions, and the "doing" does not get done either. (Drucker 1969)

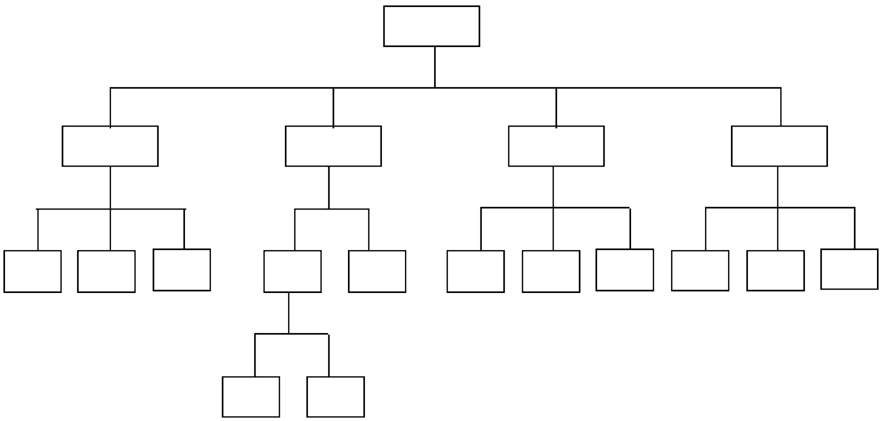


Fig. 1 A typical hierarchical organization structure that emphasizes command and control

This may derive from two basic assumptions of the industrial economy—(1) effective action is planned action, and (2) division of labor is necessary to achieve economies of scale. In time, this separation took on deeper social and psychological connotations whereby those who think, or plan were considered to be higher in status than those who did the actual work.

Before going on to discuss the advantages and disadvantages of the hierarchical organization, it is helpful to remind ourselves that design thinking as a practice emphasized values and actions that encouraged empathy, ideation, and hands-on prototyping and thus blurred the lines between thinking and doing.

3.1 Advantages of a Hierarchical Organization

A hierarchical organization displays the following advantages.

1. Efficiency: With this form of an organization. Once the task or the objective and activities are clearly planned, then the organization can actually perform those activities efficiently. The allocation of resources, which is fixed in nature, can happen faster, and the organization can respond more quickly to its environment, which presumably is also not changing. Now, as long as the environment remains static, and provided the organization is targeting the right outcome, this form of organization can be shown to be more efficient than a non-hierarchical one.
2. Legitimacy: The hierarchical organization is more familiar to most people, and a majority of people seem to prefer it. They have grown up in families where they experienced the hierarchy—father, mother, uncle, grandmother and it feels natural. Hierarchy has an intuitive feel and makes us feel safe. There is less uncertainty as individuals know their place in the organization and they know the ways of advancing within it. The familiar and ordered structure also engenders more respect for the organization. So, legitimacy and respect are tied more to a hierarchical command and control form of organization, because to some extent it has become the way we understand what it means to be organized.
3. Means of control through incentivizing: We have been socialized into hierarchies. For a majority of people who have a natural affinity to hierarchies, a hierarchical organization enables effective distribution of control through an incentive system of titles, compensation and perceived advancement that helps retain them in an organization.

3.2 Disadvantages of a Hierarchical Organization

A hierarchical organization displays the following disadvantages, which become especially acute in the context of design thinking performance.

1. Response time: If power is concentrated centrally, and there are too many levels in the hierarchy, such an organization can become very slow and find it difficult to a respond quickly in a dynamic environment.
2. Information overload: Response to changes in the environment, becomes the primary responsibility of the leaders or the people who hold concentrated power. In today's dynamic market environment, this concentration of responsibility in a small group of people becomes a bottleneck to organizational effectiveness. In a dynamic environment, people need to respond to new developments, threats, and opportunities in a much more fluid way. If everybody is passing information to the person on top and the organization is working 24 hours, it is easy to see how those who hold concentrated responsibility become overloaded with information. This then results in a management crisis whereby the amount of information and the hierarchical structure of persons no longer works as effectively.
3. Complexity: In order to better understand the relationship between complexity and organizational design, it is helpful to understand the Cynefin framework proposed by Kurtz and Snowden (2003). In it they described four domains most organizations have to navigate for strategic environmental survival:—simple, complicated, complex and chaotic. In the simple domain, relationships in the environment between cause and effect are predictable and stable, in the complicated domain, the relationship are dynamic but still stable and could be analyzed and predicted. In the complex domain, relationships are weak, emergent and unpredictable. While in the chaotic domain, relationships could not even be identified. The hierarchical organization is well-suited for domains of simple and complicated. But once one goes to the complex domain where the relationships are weak, and there are no strong signals, there's no single decision that can actually drive action. One needs to conduct hundreds of different experiments to actually figure out the right set of responses. In this domain, and more so in the chaotic domain, the hierarchical command and control organization becomes a hindrance since it constrains or restricts experimentation.
4. Learned Helplessness: Most people have relied on their hierarchy to help them resolve conflicts. For example, if two employees have a conflict and go to their boss to settle their differences then the need to engage each other, and work through such differences is minimized. Thus, a hierarchical organization indirectly reduces self-reliance and self-driven engagement to initiate action. In its extreme form this can become a pervasive quality and result in learned helplessness (Maier and Seligman 1976).

4 The Innovation Organization

4.1 Requirements

In order to arrive at the new form of organization, it is helpful to list the requirements that could enable design teams to achieve high innovation performance. Innovation activity by its nature occurs in the ambiguous and complex domain. It is by nature an emergent activity that cannot be planned ahead of time and then executed. There is trial and error. There is wayfaring. The market itself is changing. The technology is changing. Customers bring in their own adaptation to products released in the market. Therefore, performance is determined by how a team and organization responds to this dynamic environment. Dealing with a dynamic environment requires a form of organization which has some or all of the following characteristics.

1. Non-permission based: This refers to the ability to access resources quickly, without having to take permissions up and down a hierarchical chain or from a large number of people. The requirement to take permissions is an impediment to innovation. Whether that be a popular vote or whether that be an authorization from a single person. The organizational form we are suggesting should not have the requirement to take permission.
2. Resilient: At the same time, if people don't take permission and make mistakes, the form should have the ability to forgive, learn and reform. This requirement goes hand-in-hand with the ability to be non-permission based.
3. Foresighted and playful: It is also a form of organization that enables individual dreaming, and individual imagination, to become team imagination, to become organizational imagination.
4. Rapid and distributed sensing: This refers to the ability to coordinate with a large number of people outside to gather data rapidly and make sense of the data. It is a form that should enable this rapid sense-making because when in a dynamic environment, the organization's response rate depends, to a large extent, on its perception and sense-making capability.
5. Flexible Acting: The organization should leave room for a wide range of values, cognitive dispositions, and action tendencies. We need the organization to allow forms within it that don't quite fit the dominant culture to have time to grow and demonstrate value. This diversity is necessary for creating an environment where alternative ways of acting are generated, experimented with and, if proven useful, allowed to influence and change the organization.

With these requirements, we began to see that the new form of the organization that we propose is more like a flow field than a physical structure. Borrowing from the principles of fluid mechanics to construct order in such an organization, we understand that flow is directed by two aspects—the cohesive forces between the fluid molecules, and the contours of its container. Accordingly, we define the new structure by two key aspects.

1. The interactions between the participants of this organization.
2. The shape of the environment in which the "flow organization" exists.

4.2 *Reframing the Organization*

Innovative organizations need to take on more responsibility than traditional organizations. The requirements described previously point to a shift in the framing of organizations. From a mechanical clock metaphor to flow metaphor. The following factors are particularly salient when reframing organizations from hierarchical to a flow form.

Shaping Environments

In a hierarchical organization, action is influenced through planning and adherence to the plan through hierarchical channels of authority. These channels are central to the effectiveness of the organization. In a flow organization, action is influenced by shaping the environment in which it is situated. Beyond this influence, the teams working in that environment are allowed freedom to wayfare (Ingold 2009). The shaping of the environment becomes central to the effectiveness of a flow organization. This can take the form of designing physical spaces in which teams operate to designing embedded narratives for the shaped environment. Figure 2 below shows the shaped environment of a studio space within Stanford University. This environment was designed carefully to elicit certain behaviors from the teams within. The artifacts in this environment, its layout and its audio/visual characteristics, all give subtle but clear indications to the teams to emphasize certain actions. This mechanism through which an environment influences action is called stigmergy and was first discovered when investigating how ants coordinate indirectly over time and space through environmental markers (Theraulaz and Bonabeau 1999). This mechanism has also been identified in human-human communication (Parunak 2005).

Fig. 2 ME310 loft, an intentionally shaped environment for graduate level design course at Stanford University.

Team as a Fundamental Unit

The other factor that distinguishes a flow organization from a hierarchical organization is that the team is now the fundamental unit of the organization, instead of the individual. This facilitates greater information processing capacity as well as greater potential for generating alternatives for innovation. The team can be considered a super-organism in comparison with the individual acting as a micro-organism.

The term 'team' is currently being co-opted in the hierarchical organization to imply a manager or a team lead working with a small number of his or her subordinates. This is not what we mean by a team for a flow organization. A team in a flow organization is a group of people working interactively and non-hierarchically to develop joint cognition, joint motivation, joint imagination and joint action.

Activating Intrinsic Motivation

A key issue that goes with dismantling levels of hierarchy and changing to a team as the unit of organization is the issue of motivation. In a hierarchical organization, the hierarchical levels of authority imbue motivation in individuals by bestowing promotions, financial incentives and projecting the fear of being fired. In contrast, in a flow organization with its shaped environments and teams, the motivation needs to be activated intrinsically. This is more challenging but also a pre-requisite for creative engagement that could result in successful innovation outcomes (Amabile 1997). This activation of intrinsic motivation requires intentional effort on part of the individuals within the teams.

Regulation

This brings us to the institutional regulatory environment. In the past, hierarchical organizations just consisted of the individual and the institution. Flow organizations have the individual, the team, and the institution. The institutional regulation which includes legal framework, financial and accounting framework, and corporate governance framework needs to be re-oriented to recognize and enable high performance teams.

Team-Tool System

Teams in flow organization would be enabled by instrumentation and tools that form an integral system we can call a Team-Tool system. While we as humans have been using various tools for work since the dawn of civilization, it was only in the mid twentieth century that the concept of system that augments humans with tools was proposed by Vannevar Bush (1945). Douglas Engelbert's implementation of the Memex proposed by Bush led to the development of key computing technologies such as hypertext, graphical user interface, the computer mouse, and videoconferencing (Engelbert 1968). These technologies have exponentially improved our capacity to coordinate and process information.

However, we need to develop new and improved set of tools and instruments that operate at the team level and not just at the individual level, so that we can augment the collaborative capacity that results in teams becoming super-organisms. We have the beginnings of such a Team-Tool system based on the Interaction Dynamics Notation (Sonalkar et al. 2016) that gives feedback on the patterns of interactions among designers in a team. But it has not yet been fully implemented. The Team-Tool system needs to also include financial tools such as stock options for employees, and these could be implemented as smart contracts using blockchain technology such that it engenders fairness and transparency within the organization (Tapscott and Tapscott 2017).

Leadership

Leadership in a flow organization functions through demonstration by self-action followed by diffusion of influence through mimesis and stigmergy. This is in sharp contrast to leadership in a hierarchical organization that functions through plan-setting or strategy-setting followed by diffusion of influence through external narratives and the channels of command and control. Once, the hierarchical structure is dissolved in a flow organization, the often-mistaken conflation between managers and leaders would be removed. In a flow organization, teams lead, not managers.

4.3 *Why Do We Propose This Form?*

We believe the new frame proposed above, will satisfy the requirements we derived. Specifically, this new form addresses (1) the change in the dynamics of the environment, which in turn effects (2) the complexity of the organization.

Table 1 shows a graph of the requirements versus the proposed design. It should go without saying that the dynamic environment, is causing us to use the Team-Tool System, which increases the rate of change in the environment, which in turn causes us to increase the performance of the human-tool system.

Table 1 Flow organization characteristics mapped against requirements for an innovation organization

Flow organization design feature / Requirement	Team as fundmental unit	Team-tool system	Leadership	Shaping environments	Activating intrinsic motivation	Regulation
Non-permission based	X		X		X	X
Forgiveness-based (Resilient)	X		X		X	X
Foresighted and playful	X		X	X		X
Rapid & distributed Sensing	X	X			X	X
Flexible Acting	X	X		X	X	X

4.4 Example of Innovation Organization

The flow organization form that we propose above is an ideal form that may not quite exist in today's world, but is an organizational form that we are working towards achieving. There may exist aspects of the flow organization today and they have existed in the past. The example below describes what could be characterized as a partial flow organization that was developed within Stanford University in the mid 1970s. The description of the organization is paraphrased from Bernard Roth's The Achievement Habit (Roth 2015, pp. 166–170).

Roth describes how the structure of the Design Division at Stanford University changed from a hierarchy to a team-based organization that has contributed to its success. The Design Division (now called as the Design Group) was part of three divisions in the department of Mechanical Engineering. Each division had a director that was appointed by the chairman of the Mechanical Engineering department. Roth describes that he noticed flaws in this hierarchical organizational structure which manifest as faculty relegating responsibility to the division director, occasional misuse of power by division directors who sometimes put personal interest ahead of division interests, and unilateral decision-making by chairs when division directors were absent. In the mid-1970s, the Design Division which had eight faculty members decided unanimously to restructure the group to operate as a flat organization without a director. This met with initial resistance from the department chair, but the structure was nevertheless established, and it continues to operate to this day.

Roth further describes the working of this flat organization thus—"*Our new structure hinged around an hour-long weekly meeting, open to all Design Division faculty and staff. The meeting had no chair-person; we simply went around the table, taking turns bringing up any issues that required the division's decision, reporting on past happenings, and announcing future events. We operated by consensus and negotiation, almost never voting on anything. There was almost no acrimony, and people treated each other with respect, collegiality, and a spirit of shared purpose and commitment.*" (Roth 2015, pp. 167). In this description, Roth outlines the working of a typical team in a flow organization.

He further mentions the transformation that this structure engendered in motivating each faculty since everyone was in charge and they all wanted to make it work. Thus, intrinsic motivation was activated, as opposed to hierarchically assigned incentives.

Leadership was activated based on this motivation. Roth mentions that those who cared for an issue took leadership in handling it. If there was nobody who cared, that issue was not considered important enough to be handled until someone wanted to have it resolved.

Moreover, organizationally the group had much greater influence in the department than before since all eight faculty members now represented the division's interests with one voice. The division did retain flexibility in its operation, so if a director was required for any specific occasion, they appointed one of the faculty members as "director for a day".

This example shows that there are ways in which a flow organization can be realized, sometimes even within a larger hierarchical organization.

5 Discussion: Organizational Transformation and the Augmented Human

In this chapter so far, we have examined the prevalent form of hierarchical organization including its advantages and disadvantages, and we have suggested an alternative form of a flow organization for the purpose of supporting design thinking teams pursuing innovation outcomes. We have examined the requirements for innovation organization, how the flow form achieves these requirements and have described an example from the design division at Stanford University. In this section, we discuss the underlying factors that need to be considered if an organization is committed to transforming from a hierarchical organization into a flow organization.

An organization by its very implication is a formal structuring of relationships, perception and action to achieve desired objectives. As humans, we have been conditioned to understand organizations through experiencing different organizations such as family, school, university, and clubs throughout our development from childhood to adulthood. To better appreciate the effect of these formational influences, it is helpful to reflect on the analysis of Michel Foucault in his book Discipline and Punish (Foucault 1977). Foucault analyzes the beginnings of the modern prison system and asserts that the carceral system as a form of discipline is not contained within prison walls, but derives from the society beyond those walls. In fact prisons resemble other institutions such as factories, schools, barracks and hospitals because the mechanisms of control, examination and classification are similar across these institutions and the institutions fulfill the same purpose of discipline through socialization and isolation. The transformation from a hierarchical to a flow organization necessitates overcoming this conditioning. This is not just a function of what we change in the structure or the environment of the organization, but also a function of how the individuals within it can be enabled to overcome their individual conditioning to become autonomous and augmented humans. This individual transformation involves the following:

Changing Perception of Power and Control

Current socializing processes appears to emphasize status seeking behaviors. Thus, there are people that could be described as high status and those that could be described as low status. This means that if they are put in organizations, they'll recreate the hierarchical organization without any further stimulus from the environment. This could mean that there are people that are high in social sensitivity, who are not really monitoring the external environment. They are not concerned with people outside their group. Whereas there are other people that could be concerned with people outside their group, and then there are some that are not concerned with

Table 2 A comparison between the rules of a production economy, and the rules of an innovation economy

Rules of the Plantation	Rules of the Rainforest
Excel at your job	Break rules and dream
Be loyal to your team	Open doors and listen
Work with those you can depend on	Trust and be trusted
Seek a competitive edge	Seek fairness, not advantage
Do the job right the first time	Experiment and iterate together
Strive for perfection	Err, fail, and persist
Return favors	Pay it forward

people but with the natural environment or with tools in their environments. Given this natural variation, one of the changes that we can expect to make is to change the relative ratio of presence of the different types of people in a team, its diversity.

Changing Perception-Action in the Ecosystem in Which the Organization Operates
In their book titled—The Rainforest: Secrets to Building the Next Silicon Valley, Hwang and Horowitt (2012) distilled the differences in the behavior of people in the Silicon Valley to a set of norms, which contrasted sharply with norms followed by entrepreneurs, investors, and other supporters elsewhere in the world. While most governments had followed the conventional wisdom of building industrial parks close to universities as a way to spark innovation in a geographical region—Hwang and Horowitt drew attention to the cognitive and affective changes required for the new knowledge economy. In Table 2 we list the differences between the norms in the form of rules: the rules of a production economy, and the rules of an innovation economy.

Flow organizations thrive when the participants perceive their ecosystem to be a rainforest and act accordingly, while hierarchical organizations thrive when participants believe their ecosystem to be a plantation and act according to the rules of the plantation.

These perceptual changes that allow what we could call an augmented human operating system are key to enabling a flow organization to develop. The factors mentioned in Sect. 4.2 such as the shaping environment, the team, intrinsic motivation, regulation, the team-tool system, and leadership are all operating on top of this underlying perceptual configuration that exists at an individual level.

6 Conclusion

We now live and work in a global environment that is more open, connected, complex, and dynamic. These changes have come about in a very short time. They have outpaced our social organization configurations that are now overwhelmed and almost unable to adapt to these changes. In the Hasso-Plattner Design Thinking

Research program, we have been observing these changes, and developing solutions to accelerate the rate of innovation in the new digital economy. The exploration led us to develop a set of requirements to be met by any organization wanting to adopt the idea of a flow organization. Going further, we developed the design of such a system and showed how it met the criteria. In the last section we took our eyes off the form of organization and instead looked at the various changes that individuals will need to make in order to participate in the new economy of the future.

References

Amabile, T. M. (1997). Motivating creativity in organizations: On doing what you love and loving what you do. *California Management Review, 40*(1), 39–58.

Bush, V. (1945). As we may think. *The Atlantic Monthly, 176*(1), 101–108.

Drucker, P. (1969). *The age of discontinuity: Guidelines to our changing society*. Somerset, NJ: Transactions Publishing.

Engelbert, D. (1968). A research center for augmenting human intellect. *Fall Joint Computer Conference FJCC 1968. San Francisco, CA* (pp. 395–410).

Foucault, M. (1977). *Discipline and punish* (A. Sheridan, trans.). New York: Pantheon.

Hwang, W., & Horowitt, G. (2012). *The rainforest: The secret to building the next Silicon Valley*. Los Altos Hills, CA: Regenwald.

Ingold, T. (2009). Against space: Place, movement, knowledge. In *Boundless worlds: An anthropological approach to movement* (pp. 29–43).

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (Vol. 2, p. 528). New York: Wiley.

Kolko, J. (2015). Design thinking comes of age. *Harvard Business Review, 93*(9), 66–71.

Köppen, E., Schmiedgen, J., Rhinow, H., & Meinel, C. (2016). Thisisdesignthinking.net: A storytelling-project. In *Design thinking research* (pp. 13–15). Cham: Springer.

Kurtz, C. F., & Snowden, D. J. (2003). The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM Systems Journal, 42*(3), 462–483.

Maier, S. F., & Seligman, M. E. (1976). Learned helplessness: Theory and evidence. *Journal of Experimental Psychology: General, 105*(1), 3.

Parunak, H. V. D. (2005, July). A survey of environments and mechanisms for human-human stigmergy. In *International workshop on environments for multi-agent systems* (pp. 163–186). Berlin: Springer.

Roth, B. (2015). *The achievement habit: Stop wishing, start doing, and take command of your life*. Harper Collins.

Saxenian, A. (1996). *Regional advantage*. Cambridge: Harvard University Press.

Schmiedgen, J., Rhinow, H., & Köppen, E. (2016). *Parts without a whole?: The current state of design thinking practice in organizations* (Vol. 97). Potsdam: Universitäts verlag Potsdam.

Sonalkar, N., Mabogunje, A., Hoster, H., & Roth, B. (2016). Developing instrumentation for design thinking team performance. In *Design thinking research* (pp. 275–289). Cham: Springer.

Tapscott, D., & Tapscott, A. (2017). How blockchain will change organizations. *MIT Sloan Management Review, 58*(2), 10.

Theraulaz, G., & Bonabeau, E. (1999). A brief history of stigmergy. *Artificial Life, 5*(2), 97–116.

Part III
Design Thinking in Practice

New Ways of Data Entry in Doctor-Patient Encounters

Matthias Wenzel, Anja Perlich, Julia P. A. von Thienen, and Christoph Meinel

Abstract Maintenance and restoration of human well-being is healthcare's central purpose. However, medical personnel's everyday work has become more and more characterized by administrative tasks, such as writing medical reports or documenting a patient's treatment. Particularly in the healthcare sector, these tasks usually entail working with different software systems on mostly traditional desktop computers. Using such machines to collect data during doctor-patient encounters presents a great challenge. The doctor wants to gather patient data as quickly and as completely as possible. On the other hand, the patient wants the doctor to empathize with him or her. Capturing data with a keyboard, using a traditional desktop computer, is cumbersome. Furthermore, this setting can create a barrier between doctor and patient. Our aim is to ease data entry in doctor-patient encounters. In this chapter, we present a software tool, Tele-Board MED, that allows recording data with the help of handwritten and spoken notes that are transformed automatically to a textual format via handwriting and speech recognition. Our software is a lightweight web application that runs in a web browser. It can be used on a multitude of hardware, especially mobile devices such as tablet computers or smartphones. In an initial user test, the digital techniques were rated as more suitable than a traditional pen and paper approach that entails follow-up content digitization.

1 Introduction

Doctor-patient encounters are like many other social situations in that they require communication. Yet, unlike many situations of informal exchange, healthcare encounters necessitate a special, highly constrained type of communication. Doctors must obtain selective information that is illness or health relevant—as comprehensively, truthfully and quickly as possible—, while patients need to be informed about

M. Wenzel (✉) · A. Perlich · J. P. A. von Thienen · C. Meinel
Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany
e-mail: Matthias.Wenzel@hpi.de; anja.perlich@hpi.de; julia.vonthienen@hpi.de; meinel@hpi.de

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_9

their health situation. In addition to such fact-oriented objectives, just as importantly, there is also the emotive side of exchange. Here, doctors may wish to convey trustworthiness or concern, while patients may need to feel soothed and understood or to experience empowerment. And as though all this were not enough, there is also the requirement of medical documentation. Whatever serviceable information doctor and patient acquire in the course of their exchange, there remains an extra step of "communicating" it to a medical documentation system.

To an increasing extent, medical documentation systems are realized digitally. This is efficient given that medical discharge letters and other official documents must be provided in a machine written format. Furthermore, at least in Germany, patients have a right to access their medical records in an electronic format (Bundesgesetz 2013). Creating and using medical records therefore becomes a matter of human-computer collaboration.

A great deal is already known about patient-doctor communication, collaboration and alliance—especially since these factors are considered key determinants of treatment success (Castonguay et al. 2006). However, related questions that include digital media have received much less attention until now. How do doctor and patient communicate—how do they collaborate—with the medical documentation system? Taking the first steps towards an answer, we can begin with the observation that, at least at present, it is solely doctors who engage with the medical documentation system. Furthermore, the process is often cumbersome. Doctors still regularly take handwritten notes during the treatment session so as not to forget what they later want to convey in the digital documentation system. Since the system will not accept handwritten notes, everything needs to be retyped on a computer keyboard later on.

To develop medical documentation systems that suit user needs more than present-day solutions do, design thinking appears to be a highly serviceable approach.

For digital media, such as medical documentation systems, design thinking also introduces a useful frame of analysis developed by Byron Reeves and Clifford Nass (1996), called the *media equation*. Their basic idea is that people react to digital media (such as computers or internet platforms) as though they were real people or real places. This idea can first be applied in design analysis to pose the question: What if a human did what the digital media does?

In reality, however, if a human behaved like the typical medical documentation system in a practice, the conduct would seem presumptuous. Let's imagine such a scenario with a doctor (D), patient (P) and the medical record stored on a digital documentation system (M). Let us imagine the doctor and patient as they are collecting important information orally.

D: This information is important, *M*, we need to archive it.
M: I don't accept oral information. You must type it on my keyboard.
D: But that would disturb the treatment flow!? Okay, I'll take handwritten notes for now. After the session I shall retype everything on your keyboard.
P: In the meantime, *M*, could you check when I had the last vaccination?
M: I don't communicate with you, only with the doctor.

P: But you are my medical record, aren't you?
M: I don't communicate with you.

Such frames of analysis help to elucidate user needs and fuel ideation when developing future tools. We have used them in the Tele-Board MED project to rethink medical documentation.

In the past, we focused on the role of a medical documentation system during the treatment session, where it could play a more supportive role throughout. It can display information to patient and doctor, helping them keep track of the information that is already recorded as well as gaps that still need to be closed (von Thienen et al. 2016; Perlich and Meinel 2016).

We have also focused on the information output after treatment sessions, where the medical documentation system could be highly supportive again. It can create printouts for patients to take home, thus helping them remember the content of a treatment session. Furthermore, a medical documentation system can save clinicians a lot of time by creating medical discharge letters or other official clinical documents automatically based on session notes (Perlich and Meinel, in press).

One aspect that we had not looked at closely up to now is the information input. Here, present documentation systems play a highly constricting role, requiring doctors to type on a keyboard and cutting the patient off from all interactions. Therefore, we now want to place the question on center stage:

1.1 How Might We Rethink Data Entry Solutions in Healthcare Encounters?

Subsequently, we will first review related works in the field of data entry approaches, especially those used in medical contexts (Sect. 2). We will then introduce the Tele-Board MED system (Sect. 3) and suggest MED-Pad as a novel data entry solution (Sect. 4). We will explain the implementation of MED-Pad (Sect. 5) and share results from an initial user feedback study (Sect. 6), before we conclude with visions for medical documentation in the—hopefully—near future (Sect. 7).

2 Related Work on Data Entry in Medical Encounters

Notes taken during doctor-patient encounters are an important source of data feeding into medical records. The charts created during a doctor's visit include outlines that are relevant to the patient care, e.g. present complaints, history of the present health problem, progress notes, physical examinations, diagnosis and assessment, treatment plan, and medication prescriptions. In the last decades, electronic medical records have been widely applied in patient care (Hayrinen et al. 2008). There are different approaches to handling digital documentation during doctor-patient conversations.

Data entry can be carried out during the visit or afterwards, and it can be done by the care provider or by dictation to clinical assistants. Shachak and Reis (2009) conducted a literature review on the impact of electronic medical records on patient-doctor communication during consultation. They conclude that the use of electronic records can have both positive and negative impacts on a doctor-patient relationship depending on the doctor's behavior. Positive impacts are associated with the so-called interpersonal style, characterized by utilizing the computer to sharing and reviewing information together with the patient. Further potential for an improved impact on patient encounters lies in replacing keyboard typing with voice recognition, handwriting recognition and touch screens for entering data (Weber 2003). Already in 1993 Lussier et al. developed a computerized patient record software for physicians with handwriting recognition using a portable pen-based computer. However, we assume that this tool was not designed to be used during the conversation with the patient. Recent research suggests a speech recognition system that automatically transcribes doctor-patient conversations (Chiu et al. 2017).

In this chapter, we investigate the use of computerized patient records in medical encounters by using a digital whiteboard and tablet computer with handwriting and voice recognition. The described setup is a prototype. A real-life implementation however would potentially imply additional considerations regarding legal data security regulations and the clinic's existing information technology infrastructures.

3 Tele-Board MED System Overview

Tele-Board MED (TBM) is our medical documentation system that currently supports information captured during treatment sessions as well as the further information output processed afterwards. In order to improve its supportive character facilitated by advanced data entry mechanisms, we give a short overview of the system and how it can help in doctor-patient communication.

Tele-Board MED serves as the foundation for this communication by providing a shared digital workspace for doctor and patient. Notes taken by the doctor during consultations can be seen by both doctor and patient at the same time. The setup of such a consultation is depicted in Fig. 1. TBM's virtual whiteboard surface is shown on a large display or projector. The doctor and patient stand or sit in front of this device and are thus able to edit and re-arrange any of the content on the screen easily.

Tele-Board MED is a web browser-based software system to support doctor-patient collaboration in medical encounters. Its main component is a digital whiteboard web application where doctor and patient can take notes, make scribbles and draw images jointly. The application can be run across different platforms on a wide variety of hardware devices as shown in Fig. 2. The system can be used on traditional desktop computers, electronic whiteboards as well as on mobile devices, such as tablet computers or smartphones. An installation of our software is not required. It is used within a web browser, which is a common application on almost any platform.

Fig. 1 Doctor-patient setup with Tele-Board MED: Captured data can be shared and edited equally on TBM's virtual whiteboard surface

Digital whiteboard artifacts, such as sticky notes, scribbles and images, created during a treatment session are stored automatically on a central server. An explicit user initiated saving is not required. The stored digital content data can be exported into different formats (e.g. images, text documents or hardcopy printouts). Tele-Board MED can thus help to facilitate medical documentation: the relevant data is already digitally recorded and can be transformed into a document format appropriate for the desired documentation purpose, for example providing the patient with a copy of treatment session notes or creating official clinical documents.

Digitally capturing data right from the beginning during a doctor-patient treatment session is a first important step for later documentation. A duplication of work can be avoided as the data need only be acquired once. In contrast, first taking analog notes would then require re-typing them into a computer in order to create the official

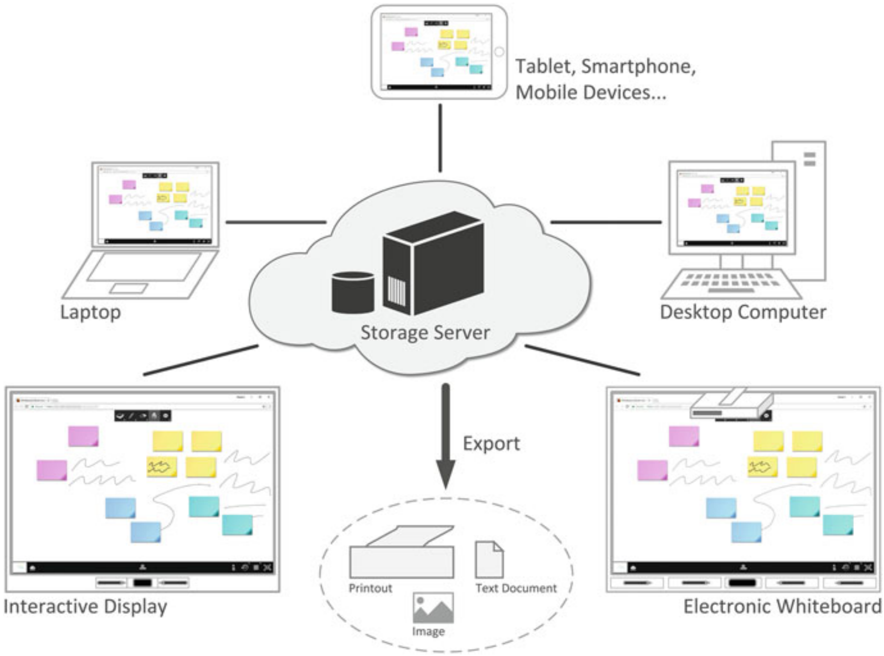


Fig. 2 Cross-platform Tele-Board MED system topology: TBM's web browser-based virtual whiteboard application can be run on different hardware devices. Jointly created whiteboard artifacts are stored on a central server automatically

document (e.g. a medical report). However, for digital text documents, both methods require the data to be in a "machine readable" text format (e.g. ASCII[1] or UTF-8[2]). The prevalent way of entering such text, even with Tele-Board MED, is via a computer keyboard. Especially during treatment sessions, this is cumbersome and disruptive for doctor-patient communication.

From our perspective, it is necessary to take a step back and focus not only on existing digital data but rather on how to capture this data. In particular, to explore methods of capturing text whereby users do not have to adapt to computers (when using a keyboard). This means developing content-capturing technologies adapted to user's natural way of working.

[1]American Standard Code for Information Interchange—character encoding standard for electronic communication.

[2]Unicode Transformation Format—8-bit—variable width character encoding standard.

Fig. 3 Schema of the MED-Pad web application on a tablet device. Hand drawn scribbles on a full-screen surface are converted into their corresponding textual representation at the user's command

4 Proposed Content-Capturing Approach in Tele-Board MED

Writing something down in order to preserve existing knowledge is a fundamental human cultural paradigm. The most common way for people to do this, and one of the first things learned in school, is to put information in a handwritten form using pen and paper. Usually, this is the easiest and fastest way to take notes. The traditional method of note-keeping is an important aspect when it comes to taking notes during doctor-patient treatment sessions.

Tele-Board MED's whiteboard application already makes it possible to scribble on the board with a pen, finger or mouse. Digital notes can therefore be created by hand. However, these digital notes are represented as vector data and not as a character encoded text format. Additionally, the whiteboard as the only input field lacks a concept that correlates to writing on a single sheet of paper.

Our goal was to enable doctors and patients to take handwritten notes digitally in a way that conveys the impression of writing on a sheet of paper. At the same time, the handwriting should be available in a text format. Our proposed solution is a web browser-based digital notepad application for handwritten notes *MED-Pad* (see Fig. 3), similar to the concept *Sticky Note Pad* described by Gumienny et al. (2011) and Gericke et al. (2012a).

MED-Pad is intended (but not limited) to be used on mobile devices such as tablet computers or smartphones. Users can draw on a full-screen surface with a stylus or finger. When pressing a button, the hand drawn scribbles are converted into their corresponding textual representation with the help of a handwriting recognition system. The recognition process has to be triggered manually. Recognizing text

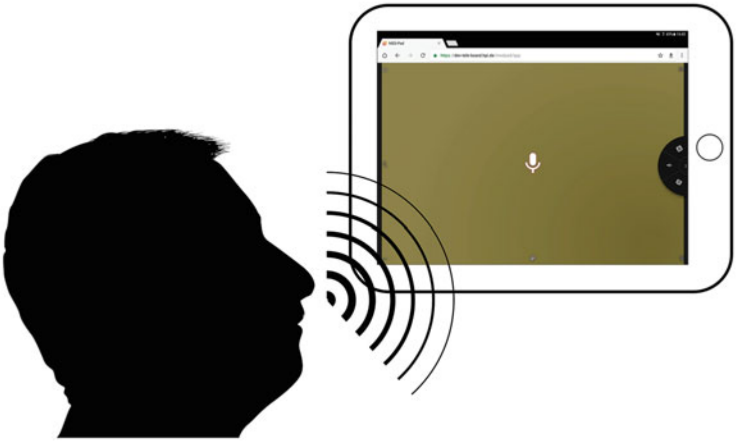


Fig. 4 User applies MED-Pad's integrated speech recognition to capture spoken words as text. Begin and end of the recognition is triggered by the user

while writing could distract users or could be even annoying when drawing visuals not intended to be recognized. However, the mechanism of how and when recognition is carried out is currently part of our further elaboration.

While focusing mainly on capturing handwritten information, we also look at new ways of content capturing. Storing spoken words as text is a technique we want to apply and investigate. For this purpose, we have also integrated speech recognition functionality in our MED-Pad application. By pressing a button, users can employ the hardware device's built in microphone to get their spoken words translated into text. This is depicted in Fig. 4. The recording has to be manually started and ended. It is intended for short notes and not to be used as a voice recorder over a longer time. Therefore, users have to think about the actual note they want to capture before starting the recognition.

MED-Pad's primary content capturing focus is the collection of textual data. Especially with regard to medical documentation, textual data is the prevalent format. This textual focus is reflected by the content capturing mechanisms described above. However, there might be situations in doctor-patient encounters when textual content capturing is not sufficient, too cumbersome or even impossible (e.g. visualizing a physical object). In such a case, a picture is worth a thousand words. Thus, MED-Pad enables users to capture images or take photos right from within the application as shown in Fig. 5.

Up to this point, all captured information is stored locally within the MED-Pad application. In order to store the captured data in our remote Tele-Board MED system, the user has to send the data to our server which in turn sends it to the corresponding whiteboard application. The user can see the textual information as sticky note within Tele-Board MED's whiteboard application (see Fig. 6). This process is perceived by the user when the sticky note "arrives" on the virtual whiteboard.

Fig. 5 When text is not enough: Using images or taking photos directly with MED-Pad

In order to integrate MED-Pad into our existing Tele-Board MED system landscape, we continue to follow our hard- and software agnostic approach. Apart from a modern web browser, MED-Pad requires neither a special operating system nor the installation of any (third party) software or proprietary web browser plugins. However, this kind of user-focused out-of-the-box usability involves some technical challenges when implementing MED-Pad.

With MED-Pad our main goal is to allow a more natural content capturing mechanism while preserving the resulting data in character encoded format. Handwriting recognition (HWR) is a complex task. Compared to Optical Character Recognition (OCR), human handwriting has a much higher amount of styles and variations. In order to be practically useful we aim to use an existing handwriting recognition system. We therefore had to elaborate on HWR and speech recognition systems in order to integrate such a system in MED-Pad.

When looking at the development of mobile applications, there is usually the choice between native (platform specific) apps and web apps. Both come with their respective pros and cons. Native apps work with the device's built-in features. These apps tend to be easier to work with and also perform faster on the device. In contrast, web apps are oftentimes easier to maintain, as they share a common code base across all platforms. Additionally, a web app does not need to be installed and is always up-to-date, since everyone is accessing the same web app via the same URL seeing the most up-to-date version of the system. However, the advantage of a web browser-based multi-platform application usable on any device comes at the price of being limited to web browser's provided resources and Application Programming Interfaces (APIs).

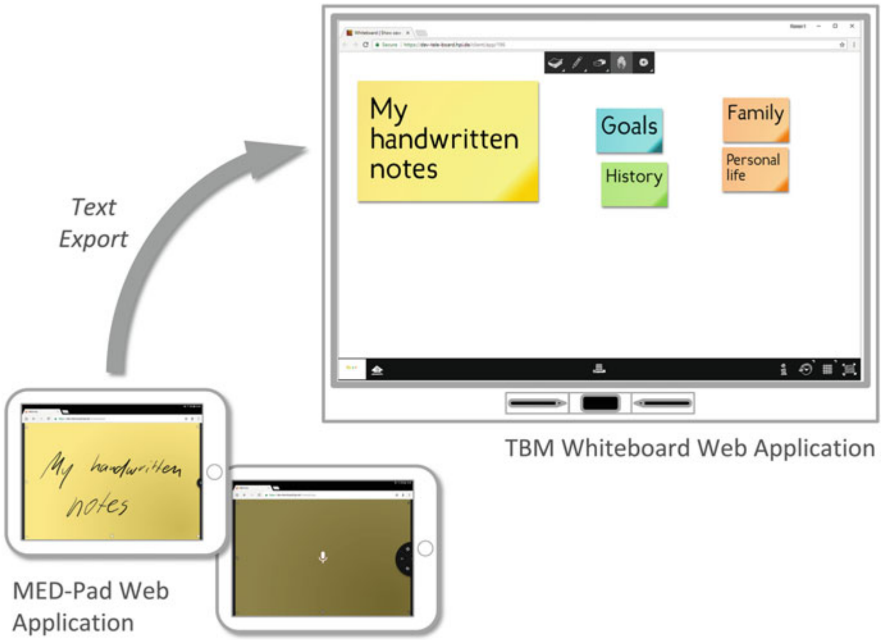


Fig. 6 Putting it all together: user's captured textual notes recognized from handwritten notes and speech with the help of MED-Pad application is sent to Tele-Board MED whiteboard web application. The recognized text is displayed on a sticky note and can be jointly re-arranged by doctor and patient

When we implemented MED-Pad, we decided on a web app-based solution. We kept following Tele-Board MED's cross-platform approach and its advantages. APIs offered by web browsers for accessing the features of a mobile device were sufficient for our first prototype implementation. This is described in more detail in the following section.

5 MED-Pad Implementation

The concepts regarding MED-Pad and its recognition features described so far, have been implemented as a web application. For this, we had to develop different components, integrated into the Tele-Board MED system architecture.

5.1 MED-Pad Components in the Tele-Board MED System Architecture

MED-Pad consists of client- and server side components as shown in Fig. 6. The web application is run inside the user's web browser and allows note taking by handwriting, speech, keyboard input and images. It can be requested just as any other URL from the Tele-Board MED website. The application assets like markup, styles and scripts are delivered by the TBM web server.

The web server also provides REST[3] endpoints for sending MED-Pad content data and for using the handwriting recognition service. This service is offered by a dedicated HWR server in the TBM network. When a user clicks the HWR button in the MED-Pad application, a respective request is sent to the TBM web server, which in turn forwards it to the HWR server. Once the recognition has finished, the result is sent back to TBM web server and then back to the MED-Pad application (see Fig. 6).

The speech recognition is also handled by a server side component. Here we rely on an external service outside of TBM's controlled scope. This is implied by the native *SpeechRecognition* API provided by the Google Chrome Browser we use in MED-Pad web application.

Access to server side MED-Pad REST endpoints is controlled by the TBM web server. Users must provide respective credentials in the MED-Pad application in order to send or receive any further content data apart from the application assets.

5.2 MED-Pad Content Capturing Web Application: A Progressive Web App

MED-Pad web application is implemented as Single Page Application (SPA) using plain JavaScript, HTML and CSS. It is a web page, once it is loaded, that completely runs in user's client side web browser. All necessary assets are retrieved with a single page load. Additionally, our application is designed as a Progressive Web App (PWA), a term introduced by Google to denote web applications relying on certain architecture and web technologies (e.g. Service Worker). PWA's main characteristics, among others, are: a native app-like look and feel, offline capabilities, and a secure origin served over an encrypted communication channel (Russell 2015, 2016).

MED-Pad's application code is cached in user's web browser making it available even if there is no internet connection. Furthermore, it is installable. For instance, on mobile devices running Chrome for Android, an app icon is added to user's home screen. Afterwards the app is started like a native app in full screen mode as shown in Fig. 8. The user's last actions (e.g. when drawing scribbles in the app or typing a

[3]Representational State Transfer (REST)—Interoperability architecture for distributed hypermedia systems.

text) are preserved in the last state automatically when the app or the web browser is closed. This way, MED-Pad can be used almost like a native mobile app.

The main interaction part of MED-Pad is its drawing surface. Technically, this is an *HTML-Canvas* element where the user can draw using a pen, a finger or a mouse. A crucial issue is that writing on the surface looks and feels as natural as possible. Thus, rendering speed and visual appearance were relevant measures during development (e.g. by using additional features for rendering, such as pressure and writing speed).

The resulting hand drawn content is stored internally as ordered vector data which builds the basis for handwriting recognition. When a user has finished taking handwritten notes, the vector data is sent to the TBM web server to be processed by the HWR server respectively. The recognition result is displayed in MED-Pad afterwards allowing the user to adjust it by means of the device's (virtual) keyboard. The character encoded text can then be sent to the TBM whiteboard application as depicted in Fig. 6.

5.3 *Handwriting Recognition as a Service: The HWR Server*

Due to the fact that MED-Pad is a web application running inside a web browser there is no client side HWR mechanism available to be utilized by our application. Hence, we came up with a server side solution similar to an approach we described in Gericke et al. (2012b) using the handwriting recognition engine called Microsoft Ink integrated in the Microsoft Windows operating system (Pittman 2007).

Handwriting recognition systems can be distinguished with regards to their input data format. Human handwriting can be digitized either by scanning a sheet of paper or (offline) by capturing strokes directly during the writing process using a special input device (online). In general, an online system handles a continuous stream of two dimensional points sent from the hardware in temporal order, whereas the offline approach uses a complete representation of the written content typically in an image format (Plamondon and Srihari 2000). Online systems usually provide higher recognition rates since these systems are able to take temporal as well as spatial information into account. Furthermore, an extraction of spatial data during an image pre-processing step can be discarded since the input is already in a vector data format.

The Microsoft system used is an online handwriting recognition. In our approach with MED-Pad we use Microsoft Ink with offline data. In our definition here the property "offline" represents the fact that the vector data is completely stored first on the client side and the recognition will not be performed during writing. The advantage of having a temporal component for recognizing text, also has the downside of being prone to stroke order variations (Prasad and Kulkarni 2010). This happens when there is a spatio-temporal mismatch within the provided vector data, which is however more likely when applied to larger areas of handwritten content (Gericke et al. 2012b). Within our proposed solution for MED-Pad we intend

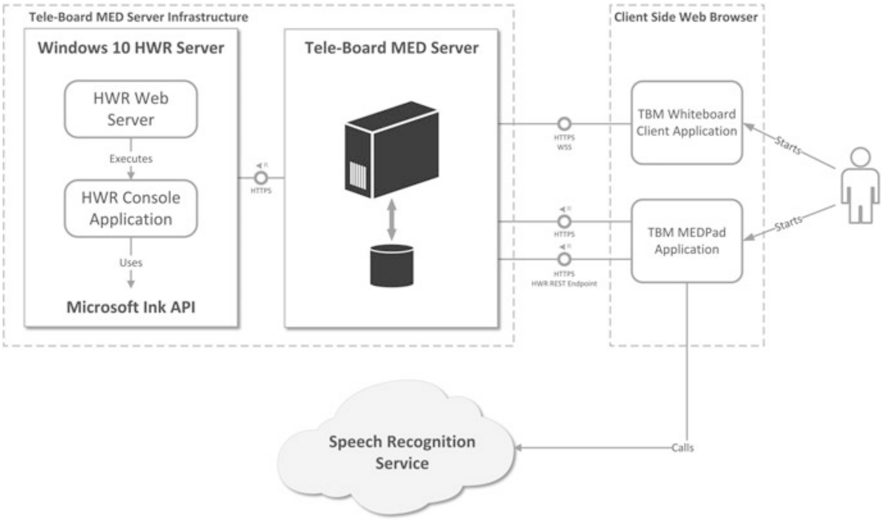


Fig. 7 Basic Tele-Board MED system architecture: TBM MED-Pad and whiteboard web applications run in client side web browser. Handwriting recognition is done by a dedicated Windows 10 server. Respective HWR requests are forwarded to this server inside TBM's server infrastructure. For speech recognition a public cloud service is used

to apply the HWR only on smaller chunks of handwritten data (i.e. at times when this aspect can be somewhat neglected).

The part of Med-Ped that facilitates handwriting recognition (see Fig. 7) is implemented as a web service and consists of three components: (1) a server machine with a Microsoft Windows 10 operating system, (2) a web server providing the service endpoint handling recognition requests from the TBM server and (3) a command line application that does the actual handwriting recognition by accessing the Windows 10 provided Microsoft Ink API.

For network communication between different machines and platforms a common format is needed. Since it is not possible to access the Microsoft Ink API directly from outside the Windows machine, we implemented a web server that handles recognition requests from outside the system over the common HTTP (S) protocol. This way, the web server defines the recognition access point. We implemented the web server using JavaScript on the basis of NodeJS.[4] Request data originating from MED-Pad web application consists of handwriting vector data in JavaScript Object Notation (JSON) format. Furthermore, the language (a two letter ISO-639-1 abbreviation, e.g., "en" for English or "de" for German) that should be used for HWR is given. This is important for the recognition engine since there is lexicon used to provide contextual knowledge in order to improve recognition rates. Currently we are supporting two languages with our system: German and English.

[4] A JavaScript run-time environment typically used server-side—https://nodejs.org/

The language is taken from the internationalization settings set in MED-Pad, i.e. the user interface language the user has selected in the MED-Pad web application. Once a recognition request arrives at the NodeJS web server the latter invokes the recognition command line application and hands in the JSON data and the requested language.

The command line application is written in C# and encapsulates the actual recognition process. At first, the incoming JSON formatted data is parsed and transferred into Microsoft Ink API required data structures, which represent the user's original handwriting strokes. Second, based on the given language an appropriate recognition object is created that processes the previously transferred data structures. As a result the recognized text is gathered by the calling NodeJS web server which sends the text back to TBM web server.

5.4 Gathering Content with Your Voice with Speech Recognition API

Our approach for transforming a user's spoken communication into its character encoded symbolic representation differs from the HWR solution. We rely on a client side, web browser available implementation of Web Speech API (Shires and Wennborg 2014). The API consists of two parts: speech synthesis and speech recognition. We are only using the recognition part within MED-Pad. Speech recognition is a highly new interface currently only provided by Google Chrome web browser.[5]

Use of the API is quite straightforward. On provided user interaction such as a button click, the browser requests access to the system's microphone. Once access is granted by the user, the system listens to the user's voice. On another click the listening stops and the recognition starts. The actual recognition is not processed on the user's system but rather on a dedicated, not specified cloud based recognition service which tends to be a Google server. As with our HWR service, speech recognition depends on server side processing but as opposed to HWR not on a TBM controlled server. Figure 9 depicts the MED-Pad application after the user has clicked on the recording button initiating the voice listening phase. Speech recognition also needs the required language to be specified. We employ the selected user interface language as the desired recognition language. As soon as the server side recognition process has finished, the textual result is displayed in the application and can also be adjusted by the user and afterwards sent to the TBM whiteboard application.

[5]https://caniuse.com/#feat=speech-recognition—Retrieved Dec. 2017.

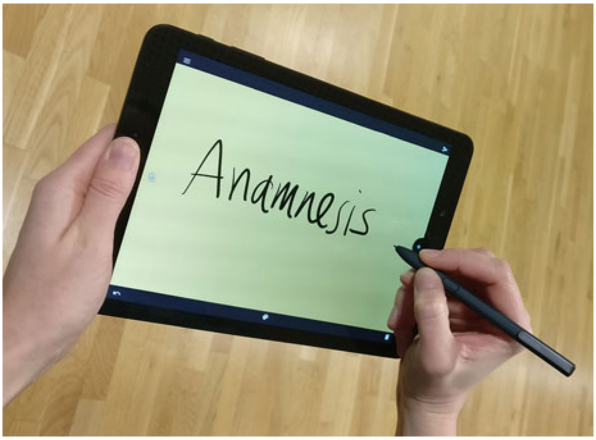


Fig. 8 Tele-Board MED-Pad application on a tablet device. Handwriting can be translated into character encoded text

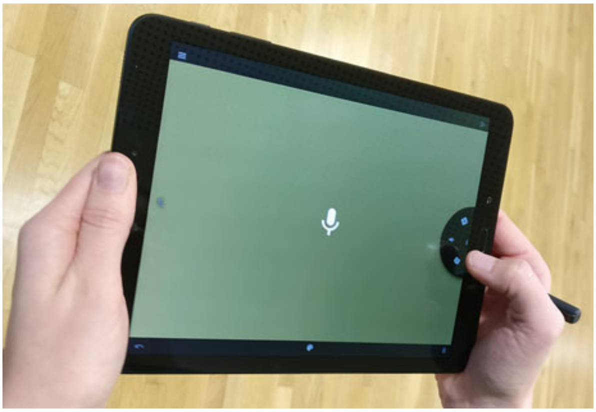


Fig. 9 Tele-Board MED-Pad application on a tablet device. The user's voice can be recorded in order to be transformed into character encoded text

5.5 Retaining Physical Artifacts with the Help of Image Capturing

There are situations when it is important to capture artifacts beyond textual information. For that reason, the MED-Pad web application allows for using pictures from the local device and annotating these pictures using scribbles or text.

We implemented this functionality by setting appropriate image MIME[6] types (e.g. "image/jpeg") to an *HTML-File-Input* element's "accept" attribute. Clicking the respective image button in MED-Pad's user interface provides different options. On traditional desktop computers a classical file upload dialog is shown, on mobile

[6]Multipurpose Internet Mail Extensions (MIME)—Standard for defining email data format. Also used for content-type definition outside of email.

devices users' are offered the option of using a local camera or selecting a photograph from their photo collection (Faulkner et al. 2017).

6 Initial User Testing

We gathered user feedback on the different methods of content capturing that are applied in a dialogue setting. Therefore, we could gain an understanding of how users feel when capturing content in the proposed ways. Two groups of two persons conducted three short interviews and were asked to apply the three approaches: (1) traditional pen and paper, (2) MED-Pad handwriting recognition with digital pen and a tablet computer, and (3) MED-Pad speech recognition with a tablet computer. We asked the dialogue partners to converse about a colloquial topic, such as their daily way to work, the perfect breakfast, and preparations for Christmas holidays. The interviewer was asked to take notes. The language of conversation and documentation was German.

We gathered qualitative feedback via free form text fields in a questionnaire and a subsequent discussion. The handwriting recognition was perceived as fast and less distractive than traditional pen and paper. One interviewer perceived it as a relief to be able to store conversation topics away by sending the captured notes to the whiteboard. However, the usage of the writing space on the digital sticky pad was not clear. While we designed the MED-Pad as a digital sticky pad intended to capture keywords or short phrases, one person captured all notes on one digital sticky note. Apparently, the writing area was rather perceived as analogous to a paper sheet instead of a small sticky note. The fewer words are captured when the recognition is triggered, the quicker the machine-readable text will appear. Up to three lines with up to four words each—the recognition happens instantly. However, over a certain number of written words, the recognition slows down and becomes unusable for our purpose of instant capturing.

The speech recognition was perceived as time-efficient and easily integrable in the conversation. However, the two modes of speaking for conversation and speaking for dictation were not easy to distinguish. Either the conversation was paused to capture premeditated content, or the conversation kept the natural flow while the recording was started alongside. With the latter approach, there is a tendency to capture long sentences, because there is no incentive to synthesize the spoken words.

Both handwriting and speech recognition results were acceptable, however contained some errors. During the course of the interview, it was not possible to quickly fix incorrectly detected words, since the text editing via the integrated digital keyboard was a bit cumbersome and too time consuming to do it parallel to the conversation. In sum, both digital techniques were rated as more suitable than the traditional pen and paper approach.

7 Conclusion

In conclusion, design thinking helps to rethink medical record experiences of patients and doctors; it inspires designs that are more user centered than conventional solutions. We use this methodology, endorsing the value of human centered design, in the development of Tele-Board MED.

In the past we had already rethought the role of medical documentation systems during treatment sessions, such as to render them more serviceable by supporting doctor-patient communication at eye level. We had also redesigned the system output, such as to provide doctors and patients with the documents they need: creating discharge letters or other official clinical documents for the doctor automatically on demand and providing the patient with printouts of session notes when requested.

In this chapter we have, in addition, focused on the information input, where standard documentation systems require a data entry procedure that is neither convenient for the patient nor the doctor. Patients are fully excluded from the process, cannot even see what the doctor records, and doctors must type on a keyboard, even though this tends to disturb patient-doctor communication.

To help the doctor communicate conveniently with the system we have implemented two additional data entry solutions. Doctors can write with a digital pen and their handwriting is automatically converted to machine readable text. Alternatively they can dictate and the content is, again, recognized automatically. With this intuitive technology, digital documentation does not disturb the treatment flow as typewriting might. Thus, doctors can document treatment content during the session digitally, rendering the documentation process as transparent towards the patient as they like. Even the patient can be given devices to enter information during the treatment session, if the doctor decides to do so.

With these developments we would like to return to the short scenario of a doctor (D), a patient (P) and the medical record, now worked on with Tele-Board MED (TBM). Let us explore what differences the redesign entails.

D: This information is important, we need to archive it. [Presses record button] allergic to penicillin [stops recording].

TBM displays "allergic to penicillin" on a sticky note on a large touch screen.

D: We are almost done now, except for the "family anamnesis" [points to TBM, where nothing is entered below the respective heading]. How about your close family members? Has anyone suffered from severe illnesses?

P: My grandmother died of breast cancer. Oh, and we have another case in the family. Shall I write that down?

D: Sure, if you want.

P [writes with a digital pen "grandmother died of breast cancer", "cousin was diagnosed with lung cancer"]

TBM displays the information on two novel sticky notes on the large touch screen, where the doctor moves them with a finger to the section "family anamnesis".

Thus, we suggest MED-Pad as a novel data entry solution, which adapts to the user's natural way of working. The system is designed as a cross-platform web application that can be run on a wide variety of hardware devices such as traditional desktop computers and common mobile devices (e.g. tablet computers and smartphones). Furthermore, writing down handwritten scribbles is possible even without internet connection.

Since the MED-Pad is platform-independent, future usage scenarios could encompass patients to use their own mobile devices to create content during doctor-patient treatment sessions. When following this "Bring Your Own Device" (BYOD) approach, patients would therefore not only be on eye level with their doctor in a social sense, but also technically. Having the application always with them on their personal device allows content capturing everywhere at every time even outside of a typical treatment session. Along the path towards even better support of user needs, design thinking has proven to be a highly fruitful methodology, which can inspire further healthcare solutions in the future.

References

Bundesgesetz. (2013). *Gesetz zur Verbesserung der Rechte von Patientinnen und Patienten*, Bonn.

Castonguay, L. G., Constantino, M. J., & Holtforth, M. G. (2006). The working alliance: Where are we and where should we go? *Psychotherapy: Theory, Research, Practice, Training, 43*, 271–279. https://doi.org/10.1037/0033-3204.43.3.271

Chiu, C.-C., Tripathi, A., Chou, K., Co, C., Jaitly, N., Jaunzeikare, D., et al. (2017). *Speech recognition for medical conversations*. Retrieved from http://arxiv.org/abs/1711.07274

Faulkner, F., Eicholz, A., Leithead, T., & Danilo, A. (2017). *HTML 5.1* 2nd Edition, W3C Recommendation, W3C. Retrieved December 2017, from https://www.w3.org/TR/2017/REC-html51-20171003/

Gericke, L., Gumienny, R., & Meinel, C. (2012a). Tele-board: Follow the traces of your design process history. In H. Platter, C. Meinel, & L. Leifer (Eds.), *Design thinking research: Studying co-creation in practice* (pp. 15–29). Berlin: Springer.

Gericke, L., Wenzel, M., Gumienny, R., Willems, C., & Meinel, C. (2012b). Handwriting recognition for a digital whiteboard collaboration platform. In: *Proceedings of the 13th international conference on Collaboration Technologies and Systems (CTS)* (pp. 226–233), Denver, CO, USA, IEEE Press.

Gumienny, R., Gericke, L., Quasthoff, M., Willems, C., & Meinel, C. (2011). Tele-board: Enabling efficient collaboration in digital design spaces. In *Proceedings of the 15th international conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 47–54).

Hayrinen, K., Saranto, K., & Nykanen, P. (2008). Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *International Journal of Medical Informatics, 77*(5), 291–304. https://doi.org/10.1016/j.ijmedinf.2007.09.001.

Lussier, Y. A., Maksud, M., Desruisseaux, B., Yale, P. P., & St-Arneault, R. (1993). PureMD: a Computerized Patient Record software for direct data entry by physicians using a keyboard-free pen-based portable computer. *Proceedings of symposium on computer applications in medical care* (pp. 261–264). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1482879

Perlich, A., & Meinel, C. (2016). Patient-provider teamwork via cooperative note taking on Tele-Board MED. In *Exploring complexity in health: An interdisciplinary systems approach* (pp. 117–121).

Perlich, A. & Meinel, C. (in press). Cooperative Note-Taking in Psychotherapy Sessions: An Evaluation of the Therapist's User Experience with Tele- Board MED. In *Proceedings of the IEEE Healthcom* 2018.

Perlich, A., von Thienen, J., Wenzel, M., & Meinel, C. (2017). Redesigning medical encounters with Tele-Board MED. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research: Taking breakthrough innovation home* (pp. 101–123). Berlin: Springer.

Pittman, J. A. (2007). Handwriting recognition: Tablet PC text input. *Computer, 40*(9), 49–54.

Plamondon, R., & Srihari, S. N. (2000). On-line and off-line handwriting recognition: A comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*(1), 63–84.

Prasad, J. R., & Kulkarni, U. (2010). Trends in handwriting recognition. In *3rd International conference on emerging trends in engineering and technology* (pp. 491–495).

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

Russell, A. (2015). *Progressive web apps: Escaping tabs without losing our soul*. Retrieved December 2017.

Russell, A. (2016). *What, exactly, makes something a progressive web app?* Retrieved December 2017.

Shachak, A., & Reis, S. (2009). The impact of electronic medical records on patient-doctor communication during consultation: A narrative literature review. *Journal of Evaluation in Clinical Practice, 15*(4), 641–649. https://doi.org/10.1111/j.1365-2753.2008.01065.x

Shires, G., & Wennborg, H. (2014). *Web speech API specification*, Editor's Draft, W3C. Retrieved December 2017, from https://w3c.github.io/speech-api/webspeechapi.html

von Thienen, J. P. A., Perlich, A., Eschrig, J., & Meinel, C. (2016). Smart documentation with Tele-Board MED. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research: Making design thinking foundational* (pp. 203–233). Berlin: Springer.

Weber, J. (2003). Tomorrow's transcription tools: What new technology means for healthcare. *Journal of AHIMA*, *74*(3), 39–43–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12914348

Design Thinking Pain Management: Tools to Improve Human-Centered Communication Between Patients and Providers

Nicholas Berte, Lauren Aquino Shluzas, Bardia Beigi, Moses Albaniel, Martin S. Angst, and David Pickham

Abstract This research explores the role of design thinking to improve pain management for patients and providers. Specifically, using a design thinking approach, we aimed to transform pain management from a unidimensional construct measured on traditional pain scales to a social transaction between patients and caregivers, through recognizing the behavioral, psychosocial, and environmental aspects of pain. To do so, we conducted a two-phase study which involved first developing a pain assessment intervention in the form of a novel Android-based pain management application. The novel application was prototyped and developed with a multidisciplinary team. This application was then tested with 10 post-operative patients and 10 registered nurses at Stanford Health Care. Our initial findings demonstrate that patients and nurses were able to communicate pain needs through the use of the novel application. Future studies will assess the concomitant changes in care delivery.

The proposed work impacts design thinking research through studying the use of technology to (1) solve a multi-dimensional problem involving complex thoughts and sensory features in individual patients; (2) improve communication and

N. Berte (✉) · M. Albaniel
Stanford Health Care, Stanford Medicine, Stanford, CA, USA
e-mail: nberte@stanfordhealthcare.org; MAlbaniel@stanfordhealthcare.org

L. A. Shluzas · B. Beigi
Stanford University School of Engineering, Stanford, CA, USA
e-mail: lauren.aquino@stanford.edu; bardia@cs.stanford.edu

M. S. Angst
Stanford University School of Medicine, Stanford, CA, USA
e-mail: ang@stanford.edu

D. Pickham
Stanford Health Care, Stanford Medicine, Stanford, CA, USA

Stanford University School of Medicine, Stanford, CA, USA
e-mail: dpickham@stanfordhealthcare.org

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_10

healthcare team performance; and (3) influence behavior change in situations requiring shared medical decision-making between patients and providers.

1 Introduction

1.1 The Problem

The subjective experience of pain has long been recognized as multidimensional (Melzack and Torgerson 1971; Clark et al. 2002; Ottestad and Angst 2013) comprising complex thoughts and feelings as well as sensory features. In adults, self-reported pain is most frequently assessed by using unidimensional scales that produce a numerical value ranging from 0 for “no pain” to 10 for “worst pain,” either by asking a patient directly for a numerical rating (numerical rating scale) or by having a patient set a mark on a 10 cm line anchored by the terms “no pain” and “worst pain” (visual analogue scale; Fig. 1) (Jensen and Karoly 2001).

Pain scales quantify the subjective experience of the person on a unidimensional scale. When properly used, they can provide valid and reliable information; however, relying on unidimensional assessment tools and self-report is fraught with limitations. **We believe that applying a design thinking approach can significantly improve the human experience of pain management for both patients and providers.**

It is recognized that the richness and complexity of the pain experience is inadequately reduced and oversimplified when rated on a unidimensional scale (Williams et al. 2000; Knotkova et al. 2004). Although methodologically convenient, self-reporting pain on unidimensional scales requires the patient to integrate qualities of the experience in unknown ways, leaving important distinctions, such as “differences between sensory-discriminative qualities, intensities, and affective discomfort confounding” (Goodenough et al. 1999).

A glaring problem with self-report is that it excludes a large number of patients because of the cognitive and communicative burden it requires (Hadjistavropoulos

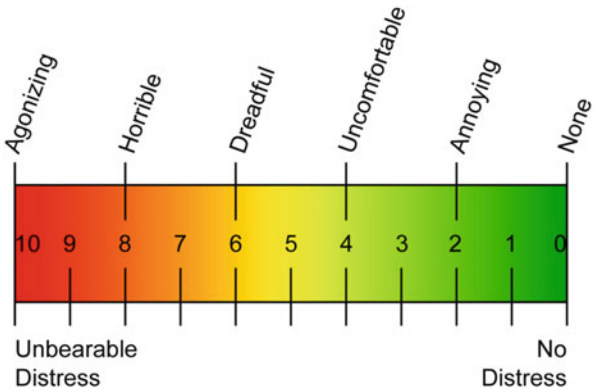


Fig. 1 Visual Analog Scale (VAS) for pain assessment

et al. 2007). Self-report requires the linguistic comprehension and social skills necessary to provide a coherent expression of pain; therefore, the strategy is problematic with some of our most vulnerable populations, the cognitively impaired (Abbey et al. 2004), the critically ill, infants, and young children (Walker and Howard 2002).

Even for people who are communicatively and cognitively competent, self-reporting pain leaves a large potential for bias and error. An inherent assumption in pain assessment is that the patient wants to minimize their pain and that the clinician wants to treat it or alleviate it. This is referred to as the "assumption of mutuality (AoM)" and unfortunately, is far from reality. Patients are often reluctant to self-report pain, and typically assume that clinicians will know they are in pain; yet, clinicians assume that patients will report pain as necessary despite this reluctance (Watt-Watson et al. 2001).

1.2 Communication Problems Between Patients and Providers

Patients provide many reasons for suppressing or masking their report of pain, including a fear of negative consequences. Patients often express concern about inconveniencing clinicians, seeming to be complaining, or having fears of tolerance or addiction to medications; and a belief that pain cannot be relieved (Ameringer et al. 2006; Cleeland et al. 1994). At the other extreme, patients might exaggerate, purposely or unwittingly, their report of pain. Reasons for exaggeration may include efforts to obtain opioids, the so-called drug seeking behaviors (Vukmir 2004), and avoiding responsibilities, or seeking compensation (Mendelson and Mendelson 2004; Mittenberg et al. 2002). A myriad of personal factors have been shown to influence or bias a clinician's response to self-reported pain. These include patients' demographics, such as age, sex, and ethnicity, as well as factors such as ***level of empathy, past exposure to pain, and personal beliefs about pain*** (Dalton et al. 1998).

1.3 Need for a Conceptual Shift

The American Pain Society introduced "**pain as the 5th vital sign"** and numerical or visual pain scales currently represent the gold-standard for assessing pain (Claassen 2005). However, to conceptualize pain as "a vital sign" implicitly assumes that it is comparable to the traditional four vital signs pulse, temperature, respiration, and blood pressure. These signs are objectively assessed, physiologically based, and easily obtained in the clinical environment. While the conceptualization of pain as a fifth vital sign highlights its importance, it is also misleading because pain is not easily measurable, nor is it an objective parameter. Pain is a subjective, multidimensional, and interactive experience that evolves over time. As such it is ***best described as a dynamic process, a transaction***.

1.4 Pain Assessment as a Social Transaction

There are compelling conceptual models that capture the complexities of the pain experience beyond its sensory dimension. An important example is the neuro-matrix model by Melzack stipulating that a wide-spread neuronal network integrates input from the body so that "experiences of one's own body have a quality of self and are imbued with affective tone and cognitive meaning" (Melzack 1989). The model stresses the importance of integrating social, environmental, and behavioral modifiers of pain (Melzack and Katz 2004). Sullivan (2008) specifically advanced the biopsychomotor model of pain, which integrates communication patterns, protective bodily behaviors, and social response behaviors. The work by Frantsve and Kerns (Frantsve and Kerns 2007) further **highlights the importance of communication in pain management in the context of shared medical decision-making**, a process that is collaborative and dynamic in nature, and is affected by demographic and situational factors from both the patient and clinician.

A theme that emerges from these models is the complexity of pain and its assessment. In viewing pain assessment as a social transaction, **pain assessment is a process, an ongoing and dynamic exchange between the patient and clinician, subject to external influences.** This relationship is one that is purposeful and goal oriented in nature, with the exchange of meaning (i.e., pain) from the patient to clinician (and back) as the essence of the transaction. However, in many clinical settings pain is still assessed as a unidimensional sensory experience, which may explain why a significant portion of patients are dissatisfied with current pain management approaches. For example, at least a third of patients undergoing surgery still report severe pain after surgery (Brennan et al. 2007).

1.5 Standard of Practice

Californian law mandates the safe practice ratio of four patients to one nurse. Nurses assess a patient's pain level routinely during 'comfort rounds.' Each hour nurses ask patients whether pain is present, and if so, the pain intensity. They address any personal hygiene issues, body comfort, and any other physical or psychological requirements. If pain is present, the nurse may provide medication or attempt non-pharmacological relief, such as distraction, repositioning, massage, or heat or cold packs. This routine is repeated over the patient's hospitalization; however, it is complicated when a patient is experiencing active pain requiring significant intervention, as the nurse's time is constrained. Currently there are no readily available communication systems available to the patient, other than a call-bell. Yet, the bell is non-specific and does not allow for remote two-way communication.

1.6 Previous Work

1.6.1 Initial Need Finding and Conceptual Prototype Development

To better understand the pain experience of patients undergoing surgery and to consider potential ways to communicate and alleviate pain more effectively, we conducted exploratory interviews with patient volunteers from Stanford's Patient and Family Advisory Council (PFAC), and shadowed nurses at Stanford Health Care (SHC) who specialize in pain management (from December 2015 to March 2016). Patients cared for by these nurses included women in labor and delivery, and patients having undergone heart transplantation.

We also conducted a literature review of non-pharmacological techniques for pain assessment and management, such as electronic gaming (Jameson et al. 2011; Kohl et al. 2013; Leanne 2012). A study showed that patients had more enjoyment, less anxiety and a greater reduction in pain with the use of active distraction (electronic gaming), than with passive distraction (television viewing) (Jameson et al. 2011). Based on the literature and our conversations with patients and providers, Table 1 provides a summary of our preliminary findings, and Fig. 2 illustrates early conceptual prototypes of interactive pain communication tools we developed. In the prototypes shown, patients are intended to squeeze the device or press a button as a way to share their pain experiences with other patients in the hospital, and to report pain to their providers.

Table 1 Features in an improved pain communication system (from patients and providers)

Must-have features	Nice-to-have features
• Improve communication between patient and provider (real time communication alerts). • Assist clinicians in making decisions about pharmacological pain relief. • Provide an easy and intuitive to use system for all patients, regardless of mobility limitations. • Recognize the complex, socio-behavioral aspects of pain.	• Provide distraction from pain (through social interaction, gaming, higher level cognitive function, etc.) • Facilitate data sharing, gathering and retrieval for patients and providers. • Provide non-pharmacological therapeutic relief through a pleasant touch and feel for patients, or way to reduce anxiety and restlessness.

Fig. 2 Early conceptual prototypes of interactive patient-to-provider and patient-to-patient communication tools

1.6.2 Pilot Study: Physical Interaction (Squeeze) as an Alternative to the Standard Pain Scale

From sharing the early conceptual prototypes with providers, we learned (as earlier work has shown), that in order for a new technology, process or intervention to be widely adopted in a health care setting, it must benefit not only the patient, but also the hospital and provider (Shluzas and Leifer 2014). As such, a human-centric, non-pharmacological pain communication system with features aimed at enhancing a patient's experience and wellbeing must likewise provide data that enables hospitals to quantitatively track pain levels and to make proper medication dosing decisions.

Since expressing pain by grip strength (cross-modality matching) is a more intuitive task than the cognitive process of assigning a numerical value (Gracely 1988), we conducted a pilot study with eight healthy subjects in the Stanford Human Pain Experimental Laboratory to determine if the magnitude of pain reported by a hand-squeezing action correlates with numerical pain reports using the standard pain rating scales. The TSA-II NeuroSensory Analyzer (Medoc Inc.) provided graded heat stimuli (up to 52 °C) to each subject's forearm. In each session, subjects quantified experienced pain on a numerical pain rating scale (Likert 1932) or by hand squeeze (dynamometer connected to a wireless data link (Vernier Systems) (Table 2). The data (unreported) showed a significant correlation between these two inputs for pain reporting.

2 Development of a Pain Management Application

A proxy measure for the performance of nursing staff is provided by the *Hospital Consumer Assessment of Healthcare Providers & Systems (HCAPHS)* questionnaire. This is provided to patients after they have been discharged from the hospital. As it relates to pain, patients are asked the following three questions:

Table 2 Summary of needs from a patient's perspective

Must have	Nice to have
Speedy method for pain assessment/reassessment	Non-pharmacological treatment recommendations: music, guided imagery, images/mindfulness
Provides human confirmation ("closes the loop")	Controls environment: dimming lights
Simplicity	Voice control options
Provides a clear plan (Expectation setting)	Personalize/tailor care options
Nurse response <2 min	Big data, learning health network
	Digital companion
	Record your story
	Comforting
	Communicate with support team (loved ones)
	Minimizes noise
	Education/informative

a. Did you need medicine for pain?
b. How often was your pain well controlled?
c. How often did the hospital staff do everything they can to help with pain?

Our prior work on pain management focused on the development of a pain *notification* device. This addressed the need for the patient to alert the provider about the presence of pain. However, it was flawed in that it did not address the providers' needs or address nurses' routine care delivery. The focus on the perception of the pain experience was important, and with the knowledge gained from the previous studies, subsequent work focused on a pain management *system* addressing both patient and provider needs.

2.1 User Insights

2.1.1 Patients

Similar to prior activities, we conducted a workshop with the PFAC ($n = 7$). First attendees were asked to think of a time when they experienced physical pain (preferably related to hospitalization). Based on this experience attendees were asked to describe how the pain experience made them feel. Emotions such as anxiety, powerlessness, vulnerability, and anger emerged. Based on these, attendees were asked what strategies they used to overcome these feelings. Remedies such as medications, loved ones, distractions, music, deep breathing, touch, and prayer were described. Once these were documented, attendees were asked to write or draw their ideal hospital-based pain experience. Two of the artifacts are shown in Fig. 3. Once an ideal state was designed, attendees described the important aspects of their

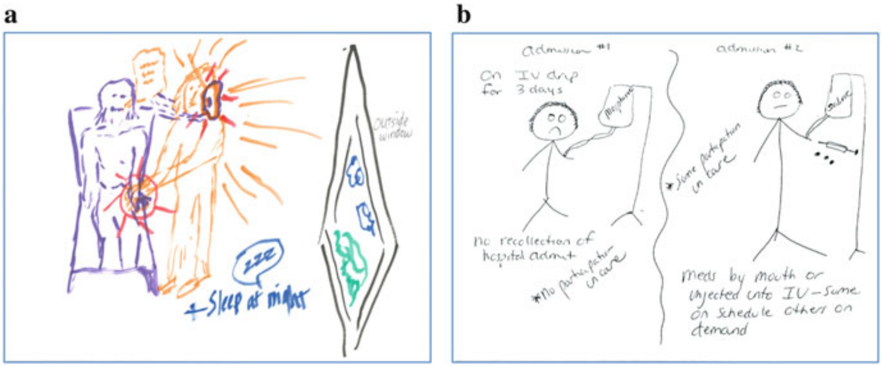


Fig. 3 (**a**–**b**) Drawings from PFAC members. (**a**) Depicts a PFAC participant laying in his hospital bed with loved ones at bedside holding his hand. An outside window is also visible letting in natural light as well as providing a view. In blue text, the PFAC member has written, "+ sleep at night" and verbalized a desire to increase the amount of sleep by reducing night time noise and interruptions from procedures such as blood draws. (**b**) Depicts a different PFAC participant at two separate admissions. Admission one depicts a sad-faced patient attached to a morphine bag. In text it sates, "One IV drip for 3 days", "No recollection of hospital admit", and "*no participation in care." Admission two depicts a patient with a natural face attached to a saline bag with pain pills and injections next to him or her. The text states, "* Some participation in care" and "meds by mouth or injected into IV—some on schedule others on demand." The PFAC participant verbalized that scheduling her medications and having "as needed medications" available provided better pain management. They also endorsed a strong correlation between their ability to participate in their care plan development and treatment and their overall care experience

experience. Such things as natural light, quietness, nurse listening, being in control, 'real' human interactions, and hands-on care were stated. To close the workshop attendees were asked to identify areas within their ideal situation that could be enhanced by technology. Attendees identified such things as pain assessment, human communication (closing the loop), screen with sounds and imagery, and individualized care.

A key insight gained from this workshop was the notion of—"Grace with technology." Patients were adamant about wanting to use technology to record their pain, and wanting more time and care from their nurses (i.e. technology that allows nurses to be more present with patients). The workshop further revealed the following must-have and nice-to-have features of a pain management application from a patient's perspective.

2.1.2 HPDTRP Focus Group

As pain is a universal phenomenon, we conducted a similar workshop to the one conducted with the PFAC, at the HPDTRP biannual workshop at Stanford University in 2017. In addition to the previous question format, we asked attendees to describe what was most important to them during their pain experience. Important to a b

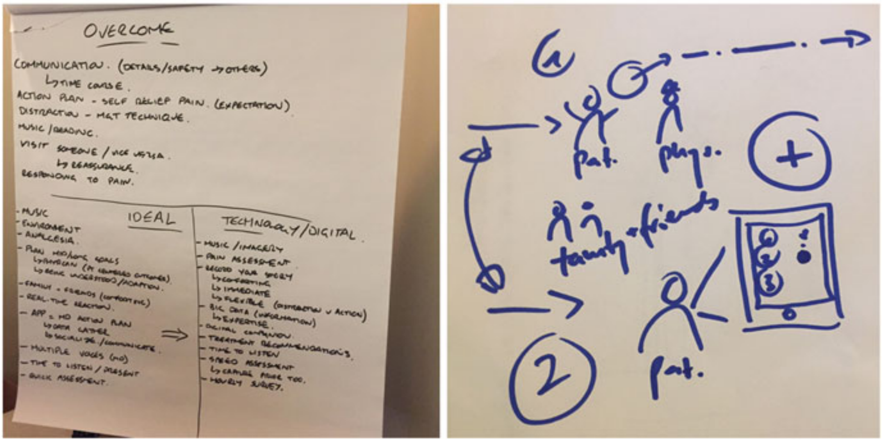


Fig. 4 Drawings from the HPDTRP community workshop. Participants were asked to think of what an ideal pain management system would offer and then describe what could be accomplished with the integration of technology. (**a**) Depicts the verbal discussion conducted during the workshop and highlights topics participants felt were important. Selections from the "Ideal" section include comments such as music, environment, pain medications, family and friends, time to listen, and quick assessment. Several of the things described under the technology and digital section are music/imagery, pain assessment, big data, time to listen, speedy assessment, and hourly survey. Participants were also asked to draw their ideal pain experience. (**b**) Depicts one participant's ideal experience. Depicted is a patient and physician interacting. Together they decide on the patients care plan. Family and friends are a key aspect of this participant's ideal experience and are depicted in the center of the image. Finally, the participant interacting with a device that allows providers and patients to communicate about their pain in real time is depicted in the lower left corner

attendees, among others, was the return of function, reduction of pain, getting empowered to fight the cause, developing contextual understanding, and establishing hope. The HPDTRP attendees provided new technical insights regarding the integration of technology into patient experience, suggesting its utility in developing individualized care through big data, ease of use with hourly pain assessment, ability to record 'my story' to improve care, and non-pharmacological support such as guided imagery and music (Fig. 4a, b).

2.2 *Providers*

2.2.1 Nursing Interviews

The lack of direct stakeholder input can limit technology transfer into the clinical setting. Because of this we wanted to ensure that nurses (one of the two main stakeholders) contribute to any future prototype. A visual non-functioning analog prototype of the proposed interface was developed and shared with nurses. The team

Table 3 Summary of needs from a nurse's perspective

Must have	Nice to have
Speedy (at a glance)	Provide remote intervention
Immediately informative	
Simple interface	
Preprogrammed response	
Accurately assess situation (pain)	

surveyed nurses at Stanford Health Care regarding their opinions on the system's overall proposed function, the display of features, and aesthetic preferences. The findings revealed that the initial prototype was in need of further development to provide the following must-have and nice-to have features (Table 3; Fig. 5).

2.2.2 Shadowing

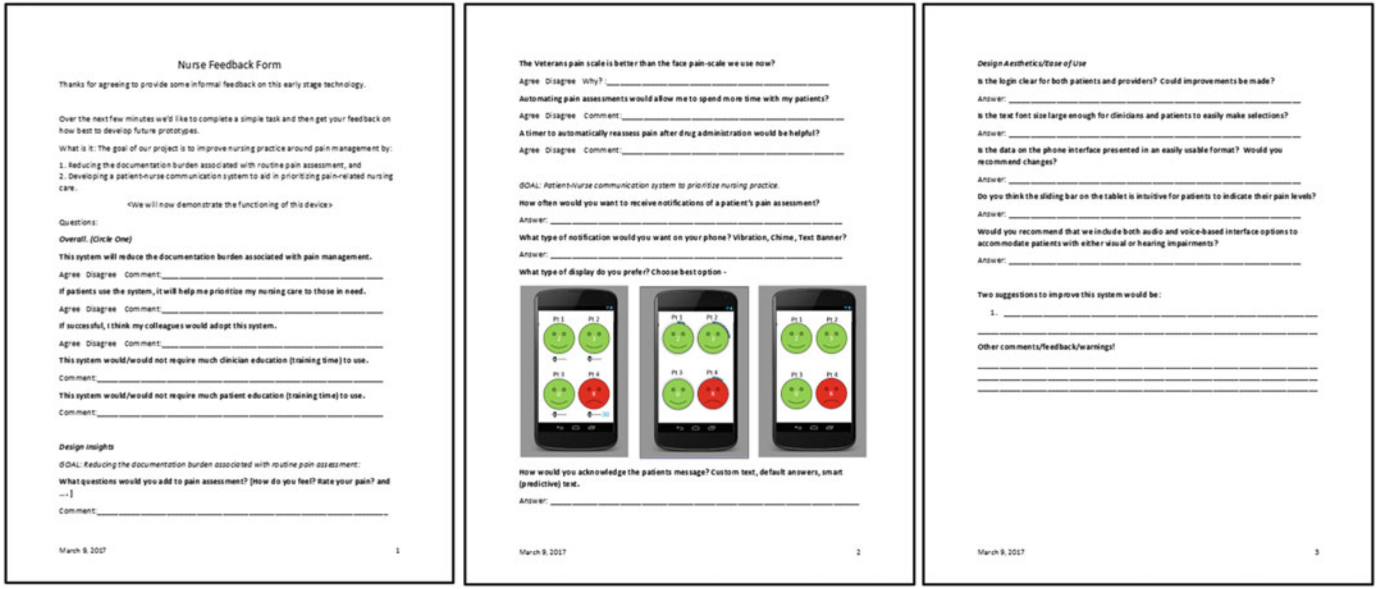
Understanding the current clinical practices was a key constraint the study team wanted to understand. This insight would help shape how the system was operationalized and drive some of the key functions. The team conducted two days of shadowing following the pain team and in a pre-operative surgical spine clinic. The pain team shadowing experience was unique in that it targeted patients with exceptionally difficult pain to manage, assess, and treat. These patients demonstrated a wide array of needs associated with pain assessment and treatment. For example, many patients seen by the pain team required high doses of opioid medications to treat their pain, other patients were known drug abusers and required more psychosocial interventions. These outliers in pain assessment highlight the need for high quality pain assessments that target each individual's unique circumstances, medical/social history, and background needs.

In a pre-operative clinic environment, patients were scheduled for elective surgery. We observed how patients were provided education on expectation setting for post-surgical pain. One key insight from these observations was that patients receive a cornucopia of information prior to surgery on a wide range of topics—questioning the retention of much this information. Hence knowledge of what to expect regarding pain and the overall experience may be dampened.

2.3 Environment

2.3.1 Direct Observations

Hospital environments are complex systems that can be broken down into subunits—typically wards or nursing units. Understanding the nursing unit's microsystem is paramount, as the rituals of the unit may differ from the established



Nurse Feedback Form

Thanks for agreeing to provide some informal feedback on this early stage technology.

Over the next few minutes we'd like to complete a simple task and then get your feedback on how best to develop future prototypes.

What is it: The goal of our project is to improve nursing practice around pain management by:

1. Reducing the documentation burden associated with routine pain assessment, and
2. Developing a patient-nurse communication system to aid in prioritizing pain-related nursing care.

<We will now demonstrate the functioning of this device>

Questions:

Overall. (Circle One)

This system will reduce the documentation burden associated with pain management.

Agree Disagree Comment:________

If patients use the system, it will help me prioritize my nursing care to those in need.

Agree Disagree Comment:________

If successful, I think my colleagues would adopt this system.

Agree Disagree Comment:________

This system would/would not require much clinician education (training time) to use.

Comment:________

This system would/would not require much patient education (training time) to use.

Comment:________

Design Insights

GOAL: Reducing the documentation burden associated with routine pain assessment:

What questions would you add to pain assessment? [How do you feel? Rate your pain? and ...]

Comment:________

March 9, 2017 1

The Veterans pain scale is better than the face pain-scale we use now?

Agree Disagree Why?:________

Automating pain assessments would allow me to spend more time with my patients?

Agree Disagree Comment:________

A timer to automatically reassess pain after drug administration would be helpful?

Agree Disagree Comment:________

GOAL: Patient-Nurse communication system to prioritize nursing practice.

How often would you want to receive notifications of a patient's pain assessment?

Answer:________

What type of notification would you want on your phone? Vibration, Chime, Text Banner?

Answer:________

What type of display do you prefer? Choose best option -

How would you acknowledge the patients message? Custom text, default answers, smart (predictive) text.

Answer:________

March 9, 2017 2

Design Aesthetics/Ease of Use

Is the login clear for both patients and providers? Could improvements be made?

Answer:________

Is the text font size large enough for clinicians and patients to easily make selections?

Answer:________

Is the data on the phone interface presented in an easily usable format? Would you recommend changes?

Answer:________

Do you think the sliding bar on the tablet is intuitive for patients to indicate their pain levels?

Answer:________

Would you recommend that we include both audio and voice-based interface options to accommodate patients with either visual or hearing impairments?

Answer:________

Two suggestions to improve this system would be:

1. ________

Other comments/feedback/warnings!

March 9, 2017 3

Fig. 5 Nurse feedback survey. This figure depicts the 20 question survey provided to nurses to assess there overall experience with early prototypes. Questions regarding their overall experience, technical design insights, the *Defense & Veterans Pain Scale* used, practice questions, design aesthetics, ease of use, suggestions to improve the system, and general comments/feedback were assessed. Each question also allowed the nurse to provide any comments or free text to elaborate on their response. These comments proved to be invaluable in the future iterations of the application. One key insight gathered and implemented was from a comment to the question, *A timer to automatically reassess pain after drug administration would be helpful?* stating, "needs to trigger in less than actual time and resend if that patient has not responded. Also, the system should alert the RN if the patient has not done the reassessment." The nursing display and system notifications were designed based off these responses policy or procedure manual. The team conducted fourteen pain management observations over a 7-week period. Each observation summary included a description of the clinical scenario, key insights from both verbal and non-verbal aspects of pain communication, as well as environmental factors associated with the assessment of pain. Factors such as noise levels, light, physical space, and overall cleanliness of the environment were all found to be factors that could affect patient's perception of pain. Figure 6 is an example of data collected at each session.

2.3.2 Review of Existing Technical Systems

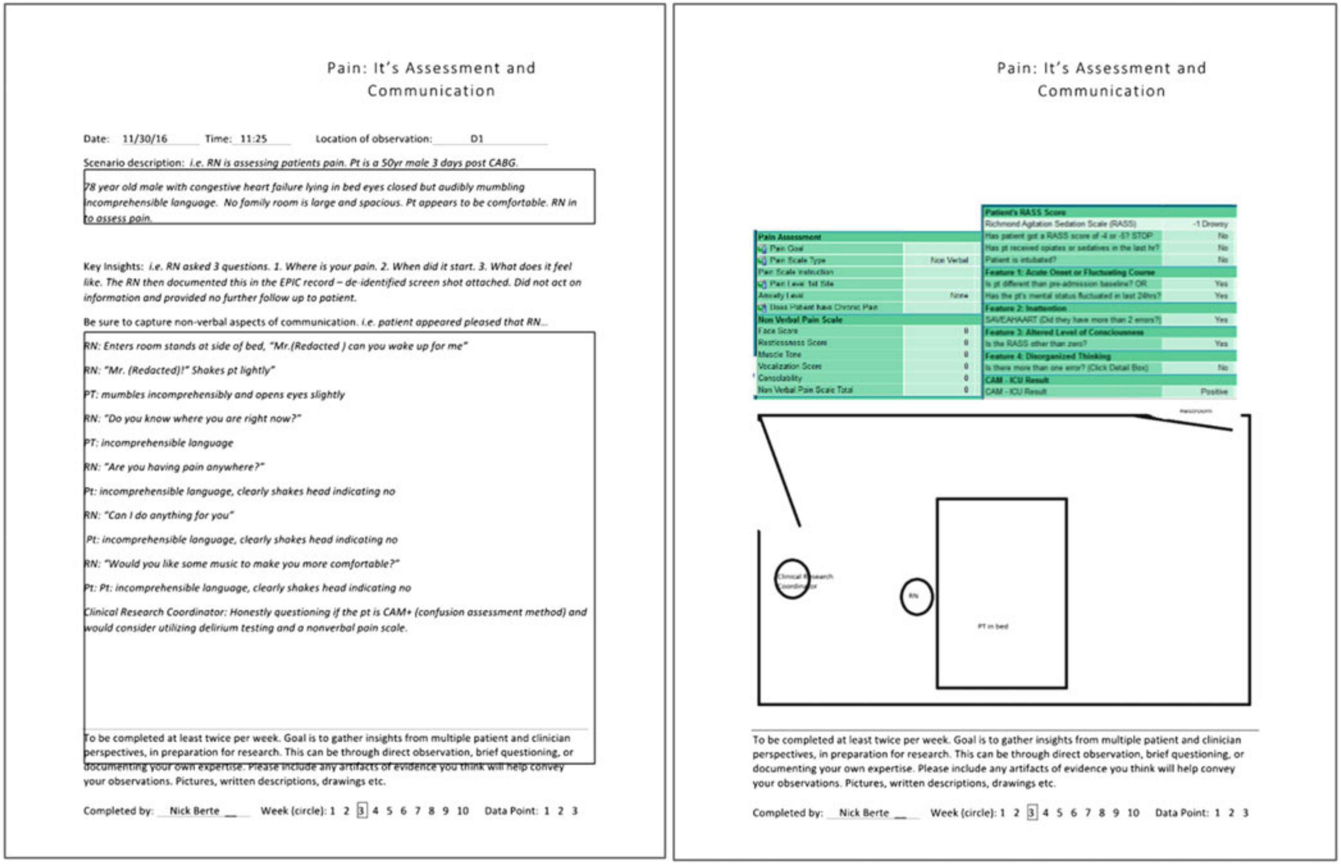
In preparation for functional prototyping, the team conducted an in-depth review of the pros and cons of nineteen commercially available personal digital assistant systems. An Android-based tablet platform was selected for patients (Samsung Galazy 9.7" Tablet), and a smart phone for nurses (Motorola G4 Play smart phone) (Fig. 7).

2.3.3 Prototype

Prototyping of the application started early to allow frequent and rapid design changes. We started with paper mock-ups of the user interface (UI) and explained the features to staff and patients. Through this hands-on prototyping we gained valuable insight into what would work and what was most valuable to each individual stakeholder. This also allowed the team to swiftly evaluate and improve multiple design iterations before coding the final application.

Once initial UI design was determined, the nursing and patient interfaces were concurrently developed. Together these two interfaces allowed the patient to communicate his or her pain needs with their respective provider. The patient interface provided a system that allowed the patient to communicate pain information, request assistance, and indicate if pain was improving.

Key feedback we received from nurses was that any solution could not 'add' to their existing documentation burden. The application offloads routine pain management documentation, providing nurses more time for direct human (patient-to-provider) interaction. Care delivery is modified as routine hourly assessments are no longer necessary. With this prototype, nurses visually identify patients experiencing pain above their preset threshold and can then prioritize his/her time to each patient. Below are screen shots from the system for both the tablet (used by patients) and the smart phone (used by nurses) (Figs. 8 and 9).



Pain: It's Assessment and Communication

Date: 11/30/16 Time: 11:25 Location of observation: D1

Scenario description: *i.e. RN is assessing patients pain. Pt is a 50yr male 3 days post CABG.*

78 year old male with congestive heart failure lying in bed eyes closed but audibly mumbling incomprehensible language. No family room is large and spacious. Pt appears to be comfortable. RN in to assess pain.

Key Insights: *i.e. RN asked 3 questions. 1. Where is your pain. 2. When did it start. 3. What does it feel like. The RN then documented this in the EPIC record – de-identified screen shot attached. Did not act on information and provided no further follow up to patient.*

Be sure to capture non-verbal aspects of communication. *i.e. patient appeared pleased that RN…*

RN: Enters room stands at side of bed, "Mr.(Redacted) can you wake up for me"

RN: "Mr. (Redacted)!" Shakes pt lightly"

PT: mumbles incomprehensibly and opens eyes slightly

RN: "Do you know where you are right now?"

PT: incomprehensible language

RN: "Are you having pain anywhere?"

Pt: incomprehensible language, clearly shakes head indicating no

RN: "Can I do anything for you"

Pt: incomprehensible language, clearly shakes head indicating no

RN: "Would you like some music to make you more comfortable?"

Pt: Pt: incomprehensible language, clearly shakes head indicating no

Clinical Research Coordinator: Honestly questioning if the pt is CAM+ (confusion assessment method) and would consider utilizing delirium testing and a nonverbal pain scale.

To be completed at least twice per week. Goal is to gather insights from multiple patient and clinician perspectives, in preparation for research. This can be through direct observation, brief questioning, or documenting your own expertise. Please include any artifacts of evidence you think will help convey your observations. Pictures, written descriptions, drawings etc.

Completed by: Nick Berte Week (circle): 1 2 3 4 5 6 7 8 9 10 Data Point: 1 2 3

Pain: It's Assessment and Communication

Pain Assessment	
Pain Goal	
Pain Scale Type	Non Verbal
Pain Scale instruction	
Pain Level 1st Site	
Anxiety Level	None
Does Patient have Chronic Pain	
Non Verbal Pain Scale	
Face Score	0
Restlessness Score	0
Muscle Tone	0
Vocalization Score	0
Consolability	0
Non Verbal Pain Scale Total	0

Patient's RASS Score	
Richmond Agitation Sedation Scale (RASS)	-1 Drowsy
Has patient got a RASS score of -4 or -5? STOP	No
Has pt received opiates or sedatives in the last hr?	No
Patient is intubated?	No
Feature 1: Acute Onset or Fluctuating Course	
Is pt different than pre-admission baseline? OR	Yes
Has the pt's mental status fluctuated in last 24hrs?	Yes
Feature 2: Inattention	
SAVEAHAART (Did they have more than 2 errors?)	Yes
Feature 3: Altered Level of Consciousness	
Is the RASS other than zero?	Yes
Feature 4: Disorganized Thinking	
Is there more than one error? (Click Detail Box)	No
CAM - ICU Result	
CAM - ICU Result	Positive

To be completed at least twice per week. Goal is to gather insights from multiple patient and clinician perspectives, in preparation for research. This can be through direct observation, brief questioning, or documenting your own expertise. Please include any artifacts of evidence you think will help convey your observations. Pictures, written descriptions, drawings etc.

Completed by: Nick Berte Week (circle): 1 2 3 4 5 6 7 8 9 10 Data Point: 1 2 3

Fig. 6 Documentation of one example of direct pain assessment observation. This two-page document was created by observing real-world pain assessments in various inpatient units. The observer would document and capture as many factors as possible including the scenario or a description of the patient, their history,

3 Summary of Study

After initial testing was completed, the application was ready for inpatient clinical testing. The study team devised a research method that would focus on (1) usability and (2) proof of concept. The study team obtained approval from the Institutional Review Board (IRB) at Stanford University. Ten post-operative patients who were cognitively intact and who could interact with the tablet interface were selected as well as the nurse caring for them. Nurses and patients received a brief training on the use of the device and then allowed time to assess its function in a real-world setting. After each patient's and nurse's interaction with the application and device, the participants were then asked to complete a digital survey evaluating their experience.

The purpose of the study was to assess the usability and perceived feasibility of an Android application for communicating a patient's post-operative pain data to nursing staff. The application was assessed from both a patient and nurse's perspective. Patients were asked to evaluate the use of the app for inputting pain information into a tablet and the nurses were asked to evaluate the ability to receive this information on a smartphone. The information collected helped identify the opportunities and barriers to using the developed application in a hospital setting and aided the research team in refining the design of an effective patient-nurse communication technology in hospitals. The use and evaluation of the application did NOT influence the delivery of medical care in any way.

Initial findings demonstrate that patients and nurses were able to effectively communicate pain needs through the use of a novel application. The study also indicated that the application could be implemented into clinical practice. Both patients and nurses indicated that this device could help to improve pain management and that it would be feasible to use.

4 Future Work

Future research aims to explore ways to integrate technology into health care to assess complex and multifaceted topics such as pain assessment. The impact that such applications may have on patient and staff satisfaction indicators is also of interest. Research may also explore ways to utilize technology, such as the pain management application, in a way that would reduce overall nursing time needed to

---

Fig. 6 (continued) and their general disposition at the time of the assessment. The observer also documented environmental factors such as loud or cramped rooms. Key insights were also captured including a script of the nurse-patient conversation. Nonverbal aspects such as tone of voice, facial expression, interruptions, and head shaking were all recorded. The observer would then capture the way the nurse documented the pain assessment in the electronic health record as well as a diagram of the layout of the room

Fig. 7 Study team actively reviewing early iterations of the pain management application. Every week the multidisciplinary study team would gather in person to review progress and assess the prototypes. This enabled early, rapid prototyping and multiple iterations to be reviewed. During this time, many technical bugs were identified as well as development of the ideal function and flow

document and assess pain. Future generations of the application could be expanded to integrate with the electronic health record in a way that could improve workflows and drive staff and nurse satisfaction. This application could also be expanded to offer a wide range of services including interventions such as music, guided imagery, meditation, or a wide array of non-pharmaceutical interventions. Incorporation of other existing systems may also greatly expand the functionality and uses of these types of application.

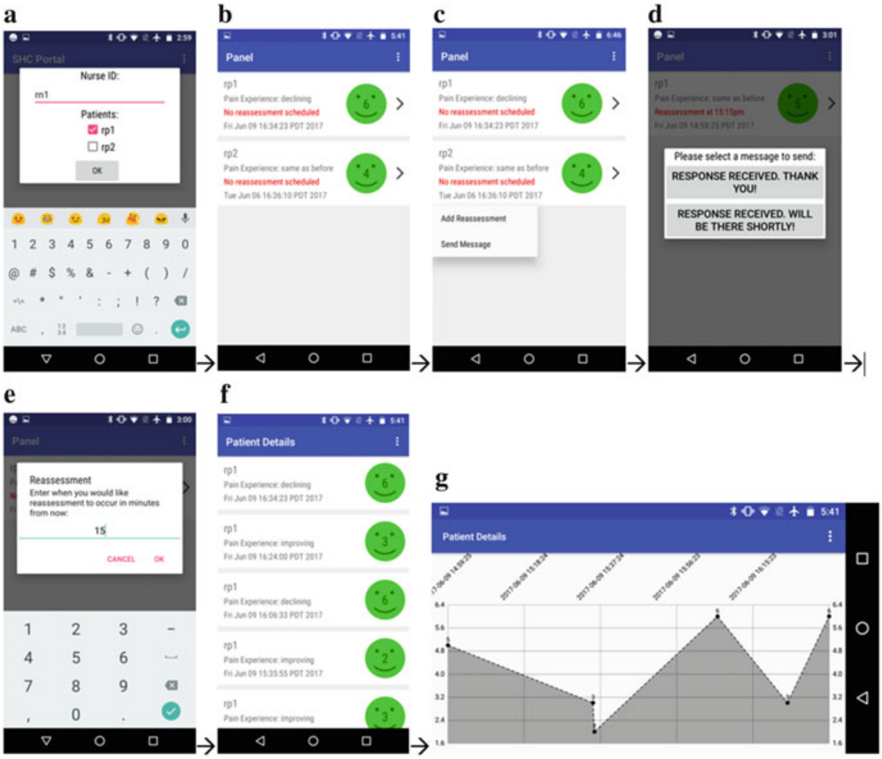


Fig. 8 Nursing user interface. Screen shots of the nursing interface are provided here sequentially *a* though *g* starting at the upper left flowing right across and then to the lower left and across. (**a**) Depicts the provider log in screen, here the provider is able to log in with their unique user ID (here shown as *rn1*). Next the provider is able to select the patients they are assigned by clicking the check box next to their patients de-identified name (here shown as *rp1 & rp2)*. After making their selection the provider then clicks the ok box and is taken to the main display (**b**). In this main display the provider is able to see the patient name, how their pain experiences is doing (improving, same as before, or declining), if there is a reassessment scheduled when their last pain assessment was, and their numerical pain number. The color of the face indicates *green* for the patient not requesting immediate assistance and *red* if the patient is requesting immediate assistance. The "nose" of the face also displays the pain number for quick reference. If the provider would like to add a reassessment or send the patient a message they can hold down on the patient tab and the two respective options will appear as seen in (**c**). If the nurse wishes to send a message, two default options depicted in (**d**) appear for the nurse to select. If the provider would like to set a reassessment, they enter it in minutes and click ok as displayed in (**e**). When the provider wishes to see the sequential assessments of a patient they can do so by single clicking on the main patient tab taking them to the individual patient documentation displayed in (**f**). Here the provider can see the trend of patient pain over time and review all of the pain assessments for that patient. In this screen the provider can also turn the device sideways to view the patient's pain data graphed over time as displayed in (**g**)

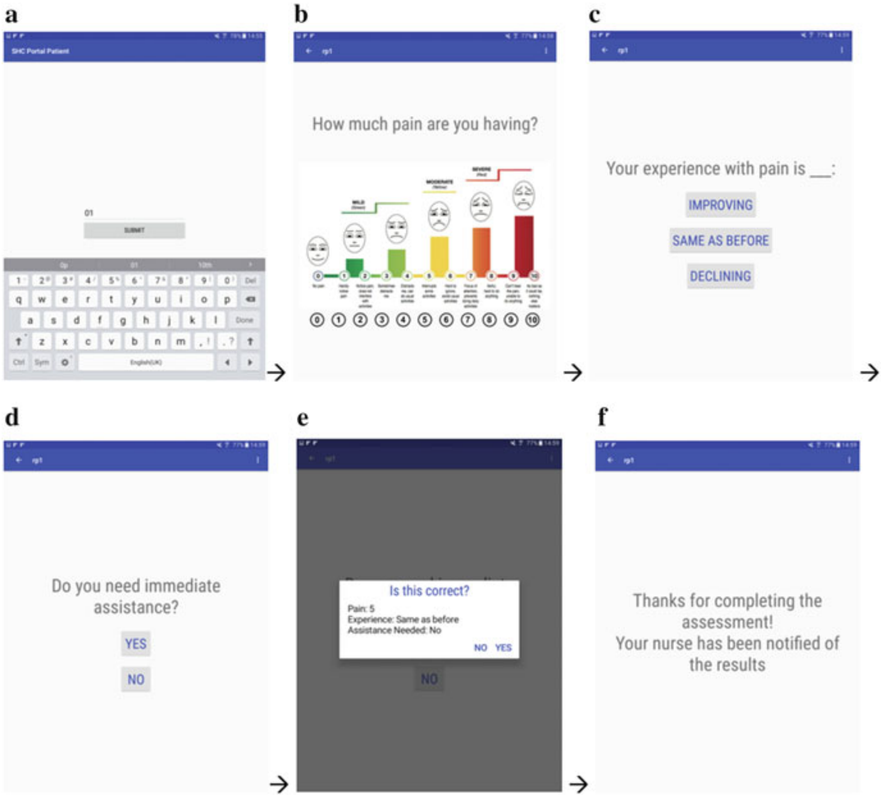


Fig. 9 Patient interface. Screen shots of the patient interface are provided here sequentially *a* though *f* starting at the upper left flowing right across and then to the lower left and across. (**a**) Depicts the patient log in screen. Here the patient will log into the application with a unique identifying name or number (for the purpose of the clinical trial, the patient was logged in using a deidentified name depicted here as 01). Patients and providers alike are only required to log in once. After log in the patient is immediately prompted to do a baseline assessment. The first screen the patient will see when doing the assessment is depicted in (**b**). Here the patient is shown the *Defense and Veterans Pain Rating Scale.* This scale displays pain levels with associated (mild, moderate, & severe) faces, bars, colors, numbers and descriptors of each number. After making a selection by touching the corresponding number to their perceived pain, the patient is then taken to question two depicted in (**c**). In this screen patients determine if their pain is improving, the same, or declining by clicking the corresponding option. Next, the patient is asked if they need immediate assistance. This can be seen in (**d**). The patient is able to select their response by touching the corresponding options. After making the selection the patient is presented with a summary of their assessment for verification as depicted in (**e**). If the patient needs to make a modification they can do so by selecting *No* and completing the assessment again. If they agree with their selections, they select *Yes* and the assessment is sent to the nurse assigned to that patient. Finally, the patient receives a confirmation screen informing them that their results were sent to the provider caring for them. When a new assessment is due the device with light up, chime, and restart the assessment. Patients can also submit an assessment at any time they need. When the nurse sends the patient a message it displays as a banners at the top of the device. The device also chimes and lights up

Acknowledgements Authors wish to acknowledge the contributions of Cecilia Cadet MS, RN and the participating staff of the C2 nursing unit, as well as the expertise provided by Barbara Mayer PhD RN, Wendy Foad MS RN, Tara Gholami MS, Mary Lough PhD RN, and John Ratliff MD.

References

Abbey, J., Piller, N., Bellis, A. D., et al. (2004). The Abbey pain scale: A 1-minute numerical indicator for people with end-stage dementia. *International Journal of Palliative Nursing, 10*, 6–13.

Ameringer, S., Serlin, R. C., Hughes, S. H., et al. (2006). Concerns about pain management among adolescents with cancer: Developing the adolescent barriers questionnaire. *Journal of Pediatric Oncology Nursing, 23*, 220–232.

Brennan, F., Carr, D. B., & Cousins, M. (2007). Pain management: A fundamental human right. *Anesthesia and Analgesia, 105*(1), 205–221.

Claassen, J. (2005). The gold standard: Not a golden standard. *BMJ, 330*, 1121.

Clark, W. C., Yang, J. C., Tsui, S., et al. (2002). Unidimensional pain rating scales: A multidimensional affect and pain survey (MAPS) analysis of what they really measure. *Pain, 98*, 241–247.

Cleeland, C., Gonin, R., Hatfield, A., et al. (1994). Pain and its treatment in outpatients with metastatic cancer. *The New England Journal of Medicine, 330*, 592–596.

Dalton, J. A., Carlson, J., Mann, J. D., et al. (1998). An examination of nursing attitudes and pain management practices. *Cancer Practice, 6*, 115–124.

Frantsve, L. M., & Kerns, R. D. (2007). Patient-provider interactions in the management of chronic pain: Current findings within the context of shared medical decision making. *Pain Medicine, 8*, 25–35.

Goodenough, B., Thomas, W., Champion, G. D., et al. (1999). Unravelling age effects and sex differences in needle pain: Ratings of sensory intensity and unpleasantness of venipuncture pain by children and their parents. *Pain, 80*, 179–190.

Gracely, R. H. (1988). Pain psychophysics. In C. R. Chapman & J. D. Loeser (Eds.), *Issues in pain measurement* (pp. 211–230). New York: Raven Press.

Hadjistavropoulos, T., Herr, K., Turk, D. C., et al. (2007). An interdisciplinary expert consensus statement on assessment of pain in older persons. *The Clinical Journal of Pain, 23*, S1–S43.

Jameson, E., Trevena, J., & Swain, N. (2011). Electronic gaming as pain distraction. *Pain Research & Management, 16*, 27–32.

Jensen, M. P., & Karoly, P. (2001). Self-report scales and procedures for assessing pain in adults. In D. Turk & R. Melzack (Eds.), *Handbook of pain assessment* (pp. 15–34). New York: The Guilford Press.

Knotkova, H., Crawford Clark, W., Mokrejs, P., et al. (2004). What do ratings on unidimensional pain and emotion scales really mean? A Multidimensional Affect and Pain Survey (MAPS) analysis of cancer patient responses. *Journal of Pain and Symptom Management, 28*, 19–27.

Kohl, A., Rief, W., & Glombiewski, J. A. (2013). Acceptance, cognitive restructuring, and distraction as coping strategies for acute pain. *The Journal of Pain, 14*, 305–315.

Leanne, J. (2012). Pain management for women in labour: An overview of systematic reviews. *Cochrane Database of Systematic Reviews, 3*, CD009234.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1–55.

Melzack, R. (1989). Phantom limbs, the self and the brain (The D.O. Hebb Memorial Lecture). *Canadian Psychologist, 30*(1), 14.

Melzack, R., & Katz, J. (2004). Pain: Psychological perspectives. In T. Hadjistavropoulos & K. D. Craig (Eds.), *The gate of control theory: Reaching for the brain* (pp. 13–34). Mahwah, NJ: Lawrence Erlbaum Associates.

Melzack, R., & Torgerson, W. S. (1971). On the language of pain. *Anesthesiology, 34*, 50–59.
Mendelson, G., & Mendelson, D. (2004). Malingering pain in the medicolegal context. *The Clinical Journal of Pain, 20*, 423–432.
Mittenberg, W., Patton, C., Canyock, E. M., et al. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology, 24*, 1094–1102.
Ottestad, E., & Angst, M. S. (2013). Nociceptive physiology. In H. C. Hemmings & T. D. Egan (Eds.), *Pharmacology and physiology for anesthesia* (pp. 235–252). Philadelphia, PA: Elsevier.
Shluzas, L. A., & Leifer, L. (2014). The insight-value-perception (iVP) model for user-centered design. *Technovation, 34*, 649–662.
Sullivan, M. J. L. (2008). Toward a biopsychomotor conceptualization of pain: Implications for research and intervention. *The Clinical Journal of Pain, 24*, 281–290.
Vukmir, R. B. (2004). Drug seeking behavior. *The American Journal of Drug and Alcohol Abuse, 30*, 551–575.
Walker, S. M., & Howard, R. F. (2002). Neonatal pain. *Pain Review, 9*, 69–79.
Watt-Watson, J., Stevens, B., Garfinkel, P., et al. (2001). Relationship between nurses' pain knowledge and pain management outcomes for their postoperative cardiac patients. *Journal of Advanced Nursing, 36*, 535–545.
Williams, A. C. C., Davies, H. T. O., & Chadury, Y. (2000). Simple pain rating scales hide complex idiosyncratic meanings. *Pain, 85*, 457–463.

InnoDev: A Software Development Methodology Integrating Design Thinking, Scrum and Lean Startup

Franziska Dobrigkeit, Danielly de Paula, and Matthias Uflacker

Abstract The debate on how to integrate Design Thinking and Lean Startup into the agile process has been addressed in the literature over the past few years. Researchers argue that Design Thinking can contribute to software development by offering support on how to understand user needs in order to derive solution and product options, whereas Lean Startup helps to learn about business and scaling strategies. Based on these viewpoints, we developed InnoDev, which is an approach that combines Design Thinking, Lean Startup and Scrum to create an agile software development process that can deliver the innovative customer-oriented products and services required by competitive companies. This study aims to describe InnoDev in detail by depicting all its phases. Our findings provide complementary perspectives regarding software development strategies, roles and techniques. This study will advance the knowledge of Design Thinking and software development by providing a detailed description of a tool that combines best practices for creating more innovative software products. The results of this investigation can help managers to evaluate their software development process in order to improve its effectiveness and create more efficient user-driven solutions.

1 Introduction

Over the years, many scholars have emphasized the importance of design for software development. For instance, Frishammar & Florén found that the early design or concept creation phases in process development are critical to the final results, and boast a large potential for cost savings (Frishammar et al. 2011). Additionally, Chin et al. focus on how to make better screening decisions for new product ideas based on the customer needs (Chin et al. 2008). Researchers suggest the integration of Design Thinking with agile software development—particularly,

F. Dobrigkeit (✉) · D. de Paula · M. Uflacker
Hasso Plattner Institute for Digital Engineering, Potsdam, Germany
e-mail: franziska.dobrigkeit@hpi.de; danielly.dePaula@hpi.de; matthias.uflacker@hpi.de

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_11

in order to improve problem understanding and solutions, as well as improving attention towards design (Lindberg et al. 2011).

Agile methods (e.g. Kanban, Scrum) have been recommended for software development due to their benefits in relation to reducing the development time, and increasing the flexibility and quality of the product (Erickson et al. 2005). However, a comprehensive systematic literature review conducted by Dybå and Dingsøyr concluded that a limitation which has repeatedly been mentioned in the literature related to Scrum is the lack of attention to design (Dingsøyr et al. 2012). These limitations can lead to severe consequences such as: a company launching the "wrong" products, resulting in poor market reception, or necessary rework requiring extra engineering hours and investments (Verganti 1997). In order to overcome these restrictions and encourage more interdisciplinary collaboration, there have been serious efforts to introduce design methods, especially Design Thinking, to IT development (Hildenbrand and Meyer 2012; Hirschfeld et al. 2011; Lindberg et al. 2011).

The debate of how to integrate Design Thinking into the agile process has been addressed in the literature in the past few years. For instance, Grossman-Kahn and Rosensweig suggest that software development teams should be guided by a clearly defined set of end goals and mindsets such as Design Thinking, agile and lean (Grossman-Kahn and Rosensweig 2012). Similarly, Hildenbrand and Meyer introduced the concept of lean thinking and developed a model using Design Thinking and agile to optimize the training experience for software professionals and their coaches (Hildenbrand and Meyer 2012). By combining the models of Grossman-Kahn and Rosensweig (2012) and Hildenbrand and Meyer (2012), de Paula developed a new model that aims to identify, implement and scale solutions in a startup environment (de Paula 2015). Häger et al. present DT@Scrum, a process model for large organizations that seamlessly integrates Design Thinking and Scrum (Häger et al. 2015).

Based on the mentioned studies, we developed InnoDev. InnoDev is an approach that combines Design Thinking, Lean Startup and Scrum in order to create an agile software development process that can deliver the innovative customer-oriented products and services required by competitive companies. This study aims to describe InnoDev in detail by depicting all its phases. This study will advance the knowledge of Design Thinking and software development by providing a detailed description of a tool that combines best practices for creating more innovative software-products. The results of this investigation can help managers to evaluate their software development process in order to improve its effectiveness and create more efficient user-driven solutions.

We consider InnoDev to be a general model applicable to different company settings (e.g. Startups, SMEs, or large organizations), however; it is necessary to validate the model with companies in order to identify whether our claim makes sense. Therefore, future research will evaluate InnoDev through an objective and systematic validation process using a combination of workshops designed to teach InnoDev and a survey designed to validate InnoDev with a large sample of software development companies.

The remainder of this article is structured as follows: Section 2 provides an overview of existing research on Design Thinking, agile and Lean Startup for software development. Our research approach is described in Sect. 3 and our general findings around InnoDev are presented in Sect. 4. Section 5 presents our evaluation protocol, and Sect. 6 closes this article with a summary.

2 Related Work

Software development has been using agile methods, such as Kanban or Scrum for several years now. These methods were developed and became popular because they often reduce the development time and increase the flexibility and quality of the product (Erickson et al. 2005). The most common approach to agile software development is Scrum (Komus 2017; Scrum Alliance 2016; Version One 2017). Scrum focuses especially on project management for projects and situations that are difficult to plan in advance, by introducing feedback loops, self-organizing teams and 1–4 week sprints as core elements (Schwaber and Beedle 2001). Another popular methodology is Design Thinking. Design Thinking has also been around for several years now and has shown to be successful as a way to develop superior products and to facilitate product appropriateness by enhancing team collaboration and improving idea generation (Beverland and Farrelly 2007). At the core of Design Thinking are four key elements: the iterative process, including various methods and tools supporting each phase; multidisciplinary teams; creative space; and a designer's mindset (Wölbling et al. 2012).

Nevertheless, both of these methods are not without shortcomings. Dybå and Dingsøyr reported a weakness repeatedly found during their comprehensive systematic literature review to be a lack of attention to design in Scrum projects (Dybå and Dingsøyr 2008). Similarly, Lindberg et al discuss that understanding the customer and finding the right solution are common among IT teams, especially in agile teams (Lindberg et al. 2011). On the other hand, Design Thinking is often criticized for not looking at the actual implementation or production of the ideas generated (Wölbling et al. 2012). Additionally, neither Design Thinking nor agile practices offer support on how to track growth and how to scale a product after its launch (Grossman-Kahn and Rosensweig 2012; Vilkki 2010).

Introducing design methods and Design Thinking to agile IT development teams is a solution proposed in literature as a way of overcoming some limitations and encouraging more interdisciplinary collaboration (Hildenbrand and Meyer 2012; Lindberg et al. 2011). As agile methods often fail to describe how requirements are gathered before the actual development, these researchers propose Design Thinking as a pre-phase to development, aimed at analyzing and eliciting requirements. This approach promises cost savings due to reductions in redesign work, as well as shortening the length of the process itself (Lindberg et al. 2011). Additionally, several authors propose including Lean Startup in such a combined methodology in order to address the issues of scaling and tracking growth (Grossman-Kahn

Table 1 Processes combining Design Thinking-and agile software development

Model	Specialty	Focus	Target group
Grossman-Kahn and Rosensweig (2012)	Lean Startup integrated and tested in a laboratory	Identify the solution + deliver a prototype	Startups
Hildenbrand and Meyer (2012)	Using lean thinking concepts throughout the development process	Identify + implement the solution	Inexperienced teams
Müller and Thoring (2012)	Combining Lean Startup and Design Thinking	Implement + scale the solution	Entrepreneurs, innovators, and startups
Häger et al. (2015)	Using Scrum to structure Design Thinking activities	Identify + implement the solution	Large software organizations
de Paula and Araújo (2016)	Integrating Lean Startup into a startup environment and testing it with students	Identify, implement + scale the solution	Startups

and Rosensweig 2012; Hildenbrand and Meyer 2012; de Paula and Araújo 2016). Lean Startup, with its build-measure-learn-lifecycle, aims at providing guidance on how to develop a product that meets its value proposition in an MVP—a Minimum Viable Product without waste (Ries 2011). Additionally, Lean Startup includes actionable metrics to assess the product performance and the user's acceptance (Maurya 2012), the business model canvas or lean canvas to develop the business side of a product and the concept of a pivot, "a special kind of change designed to test a new fundamental hypothesis about the product, business model, and engine of growth" (Ries 2011).

The characterization of new theories on how to integrate Design Thinking into the agile process has been progressing in the literature. Table 1 summarizes a selection of existing models. Grossman-Kahn and Rosensweig propose a design-led, multidisciplinary model to build innovation capacity through the integration of diverse innovation methodologies such as Design Thinking, Lean Startup and agile practices (Grossman-Kahn and Rosensweig 2012). By validating the model with a team from the Nordstrom Innovation Lab, the authors suggest that software development teams should be guided by a clearly defined set of end goals and mindsets, rather than a rigid adherence to specific tools or processes. Similarly, Hildenbrand and Meyer 2012 introduced the concept of lean thinking and developed a model using Design Thinking and agile methods to optimize the training experience for software professionals and their coaches (Hildenbrand and Meyer 2012). The authors suggest that lean thinking is closely intertwined with Design Thinking in many ways and they complement each other very well. Müller and Thoring compare Design Thinking and Lean Startup in detail, highlighting gaps, differences and intersections between the two innovation strategies (Müller and Thoring 2012). They believe Design Thinking and Lean Startup can benefit from each other since they each have features the respective other methodology is missing. As a cumulation of these thoughts they propose Lean Design Thinking, a methodology merging Design Thinking and Lean Startup. Häger et al. and Vetterli et al. present DT@Scrum, a process model for large organizations that seamlessly integrates

Design Thinking and Scrum (Häger et al. 2015; Vetterli et al. 2013). Unlike the other models, the authors use agile concepts, such as sprints and backlogs, to plan and structure the Design Thinking activities. de Paula and Araújo (de Paula 2015; de Paula and Araújo 2016) developed a model using agile, Lean Startup and Design Thinking by combining the models of Grossman-Kahn and Rosensweig (2012) and Hildenbrand and Meyer (2012). It is based on previous research (de Paula et al. 2014) and aims to identify, implement and scale solutions in a startup environment.

Although several processes and concepts that combine Design Thinking, agile practices and Lean Startup already exist, a generally accepted model has not yet emerged. Building upon the latest concepts, DT@Scrum and MOIT (which in turn were created based on some of the former concepts) de Paula and Dobrigkeit developed InnoDev (Dobrigkeit and de Paula 2017), which will be described in more detail in Sect. 4.

3 Method

In the following sections, we aim to describe each element of InnoDev in detail. To do so we will use elements that are common to method descriptions as collected by Gutzwiller. He derived five key elements as part of a method description: activities, roles, deliverables, techniques and the meta model (Gutzwiller 2013). An activity in this context describes a unit that aims to produce one or more defined results. Such activities can be structured hierarchically or in sequence. Activities are run by people or a group of people in certain roles. Roles in such cases describe a combination of activities from the view of the actor. Deliverables are the results of activities. They can also function as an input to activities and can thereby either be created or modified during activities. Techniques are tools or methods that support the creation of the deliverables. Compared to activities they are more detailed and on a smaller level. The meta model is the conceptual data model of the deliverables. The five elements and their relationships are depicted in Fig. 1.

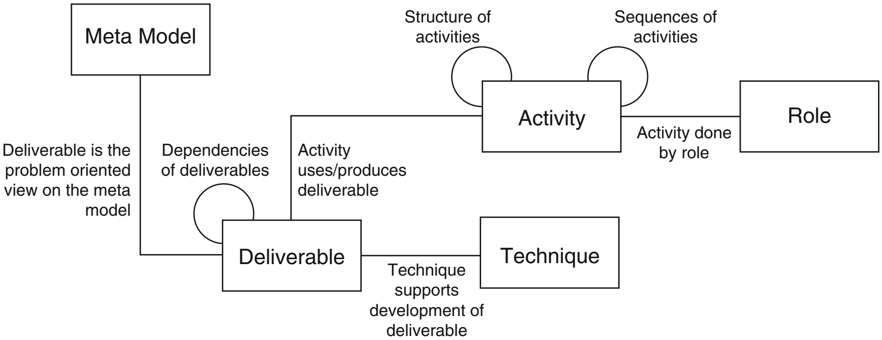


Fig. 1 Elements of a method description (translated from Gutzwiller, 1994, p. 13)

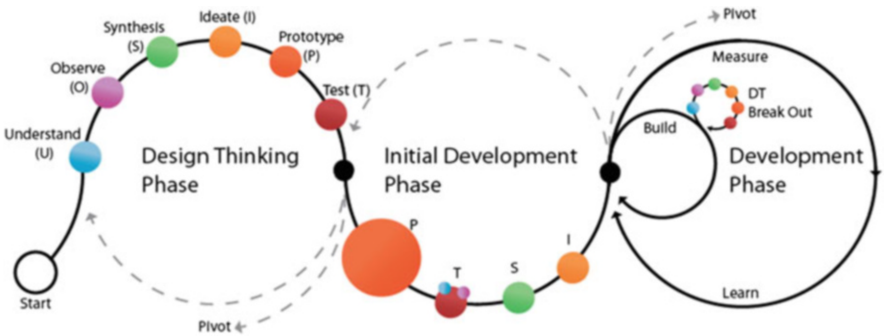


Fig. 2 InnoDev Process

4 InnoDev in Detail

Similar to former process proposals, InnoDev is based on three phases: The Design Thinking phase, the Initial Development phase and the Development phase. The process and its phases are depicted in Fig. 2.

The main difference between the three phases is the ratio between Design Thinking, Lean Startup and development activities. With an increasing understanding of the problem and the requirements for a solution, the team decreases Design Thinking activities and increases software development and business building. Lean Startup and Scrum concepts are present during all phases: each phase can be seen and implemented as a build-measure-learn-lifecycle, making use of the sprint and backlog concepts from Scrum to plan and structure the necessary activities. Thus, transparency in all activities is provided alongside flexibility to move forward with constant learning even with changing requirements or pivoting if necessary.

Before starting in the Design Thinking phase, a challenge or a general area of interest should be available to the InnoDev Team. Such a statement could be defined by the team itself or be issued by a manager or someone else outside the team. It can come in the form of a problem statement, a design brief or a simple sentence but should give the team a broad idea of the subject matter to investigate during the Design Thinking phase. Additionally, it is helpful if the team has access to potential users and stakeholders from the beginning and is sufficiently trained to use Design Thinking, Scrum and Lean Startup techniques. Armed with these pre-requisites the InnoDev Team can start the first phase of the process.

The Design Thinking phase emphasizes Design Thinking activities and is aimed at understanding user needs and related products. This phase follows the Design Thinking process as described by Wölbling et al. (2012) and Thoring and Müller (2011) to explore the problem and solution spaces and define a product vision addressing at least one of the identified problems. During this phase, Lean Startup activities support the validation of early ideas with metric-based testing and Scrum practices support project planning.

The Initial Development phase aims to refine and test the product vision from the Design Thinking phase with respect to desirability, technical feasibility and business viability ultimately arriving at a proof-of-concept prototype, following the idea of an MVP. This phase balances activities from Design Thinking, Lean Startup and software development. Business models around the concept are created and validated and ways of collecting data to monitor user acceptance are implemented. On the development side, UI as well as technical concepts are created and tested and the most important features are implemented. Throughout this phase, Design Thinking activities help to prototype, test, and refine the product vision as well as the business model and the technical concepts. Project management is done using Scrum.

In the final Development phase, the MVP will be tested and gradually extended into a full featured product according to the original concept or feedback gained during this or the previous phase. The business model and technology architecture are scaled accordingly. In this phase, the team will run agile sprints combined with lean practices to establish a build-measure-learn-lifecycle. Depending on the outcome of the learn phase the team decides whether to pivot their project or continue to the next sprint. While this phase is focused on development and scaling, InnoDev proposes to make use of Design Thinking tools in an ad-hoc manner in case of blockers or problems related to either the product or the process.

4.1 Scrum: The Overall Project Management Method Underlying All Phases of InnoDev

To structure the work during all phases of InnoDev, Scrum project management tools provide an overall framework. Scrum proposes planning work in smaller cycles of 1–4 weeks, a so-called sprint (Deemer et al. 2012; Schwaber 1997; Schwaber and Sutherland 2013). Each sprint consists of a planning, working, and reflecting on the work done and the deliverables created. All three methods that are merged in InnoDev are essentially centered around trying and learning, each using different tools and techniques. Design Thinking aims to understand and learn about problems and user needs in order to derive solutions and product options. Lean Startup aims to learn about business strategies and scaling options and Scrum aims to learn about development options and further directions during software development with changing requirements. As such, each phase of InnoDev can be considered as a build-measure-learn-lifecycle as presented by Ries (Eisenmann et al. 2012; Ries 2011)—albeit each with a different focus and therefore with a different set of actions, deliverables techniques, and roles used. Consequently, techniques to reflect on the achieved work and adjust the project accordingly are also inherent to each of the methods. In this way, Scrum is a useful method to streamline these techniques into an overall project management framework. Figure 3 depicts the core activity of the Scrum framework the sprint.

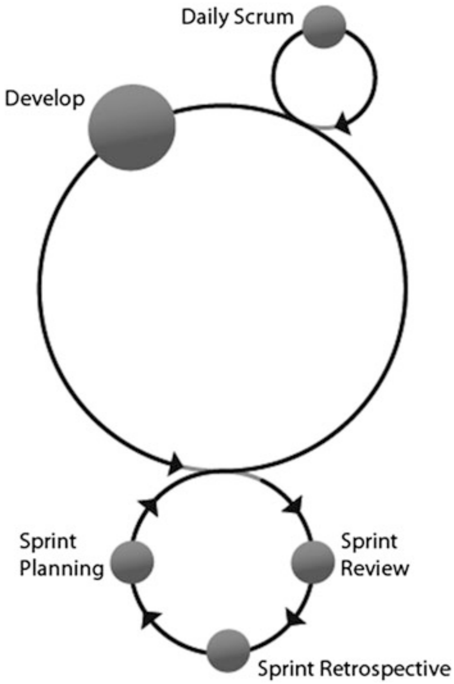


Fig. 3 Basic Scrum process

4.1.1 Activities

The core activity of the Scrum framework is a time-boxed development sprint at whose end a working version of the system under development is created. Such a sprint allows an ongoing validation of the product with the customer requirements and thereby allows the alteration of requirements if necessary. The Scrum framework proposes four meetings within one sprint: a sprint planning meeting, to decide on the work for the upcoming sprint, a review meeting, to reflect on the produced deliverables, a retrospective meeting, to reflect on the team work and process, and a short stand-up meeting, to discuss progress open work and issues in the team during each sprint. The sprint planning, the review and the retrospective should each occur once every sprint while the stand-up meeting should occur daily at the beginning of the workday.

In the context of InnoDev not all of these meetings are relevant for all phases. The sprint planning meeting and the retrospective should be held throughout the whole InnoDev process. However, retrospectives might not be necessary after every sprint. Especially during the Design Thinking phase retrospectives of the team work and process can be found as a short daily check-out. Therefore, not every sprint requires a retrospective meeting and they can be held if necessary during this phase. The review meeting is only relevant when work is split between team members. However, when working in the Design Thinking phase a lot of work is done with the whole team, thus working in sub-teams requires an immediate review of the deliverables produced. Therefore, the review meeting is unnecessary during this phase. The daily stand-up meeting on the other hand already exists in some implementations of Design Thinking as a so-called team check-in. This meeting is also known to Lean Startup experts, as they use agile development practices.

4.1.2 Roles

Scrum proposes three main roles, the Product Owner, the Scrum Master, and the Development Team. The Product Owners is responsible for collecting requirements from users and other stakeholders of the system. He also transfers them into small, understandable and implementable pieces (often in the form of agile user stories).

The Scrum Master is a specially trained moderator and coach, supporting the product owner and the development team by moderating the meetings and solving problems that occur along the way. He also makes sure everyone adheres to the rules of Scrum.

The Development Team takes care of the work items planned for each sprint and implements the selected requirements.

Again, not all roles are relevant to each phase of InnoDev. The role of the Scrum Master makes sense throughout the InnoDev process. However, as several methods and techniques play a role in InnoDev this role could be merged with supporting roles from the other methodologies such as a Design Thinking coach. The role of the product owner only makes sense once a product vision exists. Before that each member of the team should partake in eliciting and prioritizing requirements. Because in the earlier phases of InnoDev "development team" might be a misleading title, we will use only the term "team" or "InnoDev Team" instead.

4.1.3 Deliverables

The main deliverables of the Scrum framework are the Product Backlog—a collection of work items necessary to complete the project, the Sprint Backlog—the collection of work items due during the current sprint and the working software increment—the outcome of a sprint.

Scrum was developed to create software and therefore the wording of the deliverables matches this context. For InnoDev, we propose a more general wording that fits for all three phases of our process. The Product Backlog becomes the Project Backlog which will be filled with Design Thinking tasks in the beginning and only later will include software requirements. The name Sprint Backlog is still valid; however, the contents of this backlog change according to the Project Backlog. The working software increment simply becomes an increment in InnoDev. It can describe any form of progress such as user needs or solution concepts during the early phase of InnoDev or actual software prototypes and working software in the later phases.

4.1.4 Techniques

Scrum itself does not propose specific techniques. However, a number of techniques to support Scrum activities have been developed and reported. Interesting for InnoDev are the collections of retrospective games by Kerth, Derby et al., and Kua (Derby et al. 2006; Kerth 2000; Kua 2013) (which are similar to various Design Thinking techniques) and planning techniques, such as planning poker or bucket planning, which allow for quick planning in a team (Grenning 2002). Furthermore, the use of a task or Scrum board makes sense in tracking the current activities and progress in an easy and flexible way.

4.2 Design Thinking Phase

The Design Thinking phase, as depicted in Fig. 4, mainly uses Design Thinking techniques to find and explore the projects' problem statement and the solution space. Additionally, Design Thinking and Lean Startup techniques are used to validate first solution concepts. The goal of this phase is to come up with a clear product vision and corresponding low-resolution prototypes and user stories.

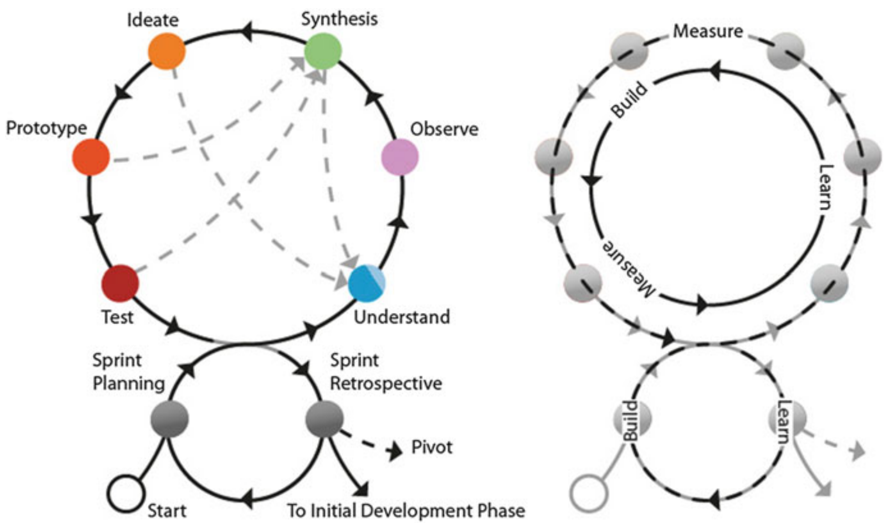


Fig. 4 Overview of the Design Thinking phase and representation of the build measure and learn concepts within this phase

4.2.1 Activities

The main activity during this mode is the Design Thinking process as described by Wölbling et al. (2012) or Thoring and Müller (2011). The team starts out with a general problem statement or an area of research and uses the initial Understand phase to collect information about the projects goals, constraints, and environment. In the following Observe phase, the problem domain is further investigated by looking at existing solutions, and getting in contact with real users and stakeholders. In the Synthesis or Point of View phase, the team reviews all the information gained from the first two phases and aims to condense them into their Point of View on the problem. In the following phase, the team uses the condensed problem statement as a basis for ideating possible solution ideas. The team reviews the generated ideas and selects promising candidates to prototype during the prototyping phase. The prototypes thus created will undergo qualitative and metric-based testing with end-users. The results of these tests will be synthesized again and, depending on the outcome of that synthesis, the team can decide to either iterate on the solution to refine it, continue with other solution ideas, or pivot and go back to "understanding" and "observing" to get a new view of the problem.

4.2.2 Techniques

Design Thinking and Lean Startup and Scrum each come with their own sets of techniques, which are useful in understanding the project environment, stakeholders, users, problem space, and solution space. Useful techniques from the Design Thinking toolbox include: 360° research, observation and interview techniques, storytelling, synthesis techniques, brainstorming techniques, various forms of prototypes, and testing techniques. This set is complimented by metric-based measurement techniques from the Lean Startup toolbox and planning techniques from the Scrum toolbox. In the following, we give a short description of these techniques and their purpose. As the number of techniques is large and still growing, the following descriptions should be considered examples and not the only usable techniques during this phase.

360° research is essentially a desk and internet research, which allows the team to quickly become well-versed in a new topic. Observation and interview techniques enable the team to get in contact with stakeholders and end-users to understand their views on the topic and problem. They can also get an understanding of their needs and pains. These techniques include: shadowing, observing participants, interviewing groups and individuals, and seeking out extreme users (d.school Stanford, n.d.). Storytelling is a great technique to share information gathered during interviews, observation or testing within the team (d.school Stanford, n.d.). Synthesis techniques aim to capture information gathered by the team and arrange it in a form that provides an overview or organizes it in a way that makes it possible to convey or derive new findings. Examples of such tools are Stakeholder maps

(Freeman 2010), Personas, Point of View Statements, a 2-by-2 Matrix or Venn-Diagrams. Brainstorming techniques are used to generate solution ideas for discovered needs or personas. They include techniques such as, hot potato brainstorming, silent brainstorming or body storming.

Prototyping techniques can be used to understand the user and the problem statement or to validate ideas. During this phase prototypes will mostly be rough and quick and easy to build for example cardboard or paper prototypes, sketches of user interfaces, or even role plays. Naturally such prototypes need to be tested with actual users to get feedback on the current solution idea and discover flaws and further needs. Testing techniques include observing users while they are trying everything out and then interviewing them afterwards using testing protocols. Furthermore, landing pages are a good way to test the user's interest in an idea. This is done by creating a website describing the future product and either tracking how and when people find the page, or adding a subscription form to actually see how many people subscribe.

4.2.3 Roles

The InnoDev Team is responsible for planning and executing the sprints during the Design Thinking phase. Such a team should consist of people from different areas of expertise, e.g. accounting, sales people, UI designers, developers or consultants, depending on what knowledge will be relevant for the project.

Potential users and stakeholders provide insights on the topic and their problems and give feedback on ideas prototypes and the project in general. For that purpose, potential users are interviewed and observed by the InnoDev Team. Potential users originate from a broad range of users in the beginning of this phase and then gradually narrow down to a target user group. The InnoDev Team tries to secure people from this group for continuous testing and feedback cycles.

For most projects, a person or group of people has formulated the original challenge. These people can be external customers or partners, or internal project sponsors (e.g., customer representatives or managers). Either way the people in this role, whom we call project sponsors, are the first contacts the team has as a way to reaching experts dealing with their challenge or topic. Thus, the person in this role is responsible for providing initial material (e.g., reports on former projects or products, market or other research that was already done or a general introduction to the topic of interest), and initial interview partners (e.g., knowledge experts inside the company or possible customers). Additionally, those in this role connect the team with other departments inside the company to enable synergistic effects and avoid duplicate efforts. Furthermore, who serves in this role provides feedback in the same way potential or target users do.

The InnoDev Facilitator can be one or more people who support the InnoDev Team. They help the team navigate through the process by introducing useful techniques, helping each role to understand what to do and how to do it, and helping the team to solve problems that arise along the way. Additionally, whoever serves in this role is responsible for moderating team meetings and discussion, as well as watching team dynamics. As such, the person or people in this role should have a solid understanding of InnoDev and its components Design Thinking, Lean Startup and Scrum in order to provide useful techniques and guidance at proper times.

4.2.4 Deliverables

The main output of this phase is a clear product vision, which will be further tested and refined in the Initial Development phase. Along the way in this phase the InnoDev Team will produce knowledge that should be documented in quick and easy ways, thus making it possible to trace ideas and decision. Such lightweight documentation could include interview summaries, collecting the main insights; filled out synthesis frameworks such as personas or matrices and diagrams; idea sheets documenting the core concepts of promising ideas, various low-resolution prototypes as well as one or more sophisticated solution prototype. These materials should make clear why each aspect of the product vision and the solution prototype have been designed the way they have. In order to make the step into the Initial Development phase, the InnoDev Team will create high-level user stories and a list of non-functional requirements based upon the materials created, the product vision and the solution prototype.

4.3 Initial Development Phase

The main goal of the Initial Development phase as shown in Fig. 5 is to create a minimum viable product based on the solution prototype and the product vision created in the Design Thinking phase. To that end, the solution will be further refined with a special focus on ensuring viability, feasibility and desirability through further exploring and testing not only the solution itself but also of possible business models and technologies to use for implementation. The outcomes of this phase will be higher resolution prototypes of the solution, the MVP, refined user stories, a list of non-functional and technical requirements and a business model.

4.3.1 Activities

The main activities of this mode are the further refinement of the solution, the validation of technical aspects and the technology to use, the creation and validation of a UX Design, the creation and validation of a business model, and the implementation of a minimum viable product, a working software system, albeit only with the most essential features and not necessarily on the final technology stack or in the final design. As such, this phase focuses on various prototyping and testing activities in the areas of software development, UX and Design and business development.

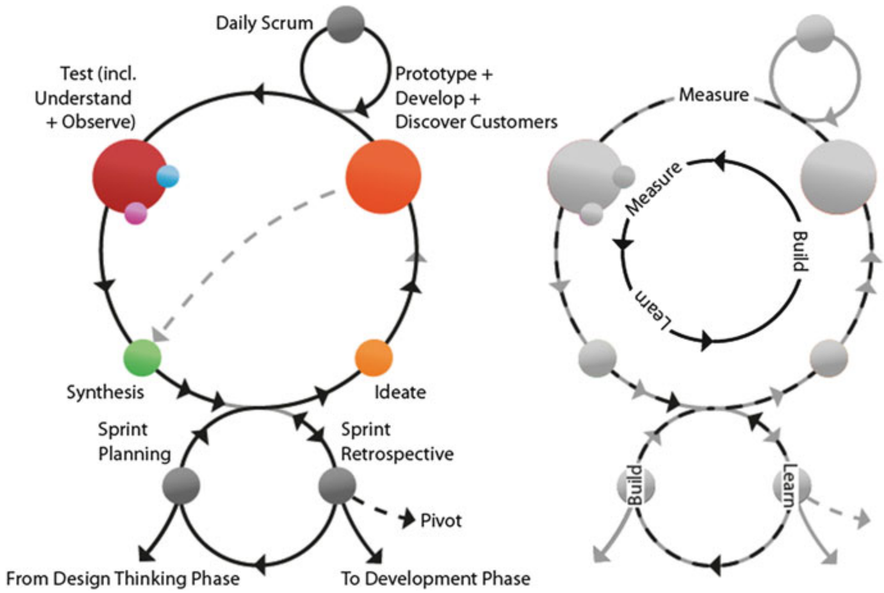


Fig. 5 Overview of the Initial Development phase and representation of the build measure and learn concepts within this phase

Initially the InnoDev Team identifies aspects of the solution prototype and the product vision that need further clarification, either in terms of detailing the concept, ensuring technical feasibility or business viability. These aspects are then refined by running through steps of the Design Thinking process as necessary (e.g., if it is unclear whether other features are needed, further interviews and observations can be initialized; if different concepts for a feature exist, these can be prototyped and tested with target users, if new features should be added, further ideation will help; or in case technical feasibility is unclear a proof-of-concept prototype can be developed. After the first refinements, and parallel to further refinements, a UI design, a software architecture, and the business model will be created and validated in further testings. Additionally, technology options are being evaluated. The knowledge gained through all these activities is then used to decide on an MVP and implement it along with further refinement of the business model and the UI and software design. Additionally, the existing user stories and requirements lists should be updated according to the new knowledge leading to more refined user stories as well as non-functional and technical requirements

4.3.2 Techniques

Core techniques during the Initial Development phase are again taken from the toolboxes of each original methodology. Design Thinking techniques used during this phase include mid-fidelity prototyping techniques and qualitative testing techniques. Lean Startup techniques include business model creation, further measurement based testings and the beginning of customer development. Scrum techniques again include project planning, and reflection and evaluation techniques. Additionally, agile development techniques created around Scrum, Lean Startup and other agile methodologies are important for implementation efforts around technical prototypes and the MVP. As in Sect. 4.3.1 the following descriptions should be considered examples and represent only a couple of the techniques that are helpful during this phase.

The most prominent mid-fidelity prototyping techniques include interactive wireframes and more sophisticated UI prototypes. Interactive wireframes can be used to evaluate interaction and navigation concepts as well as arrangements of content in the software. They can be created with simple sketches on paper adding interactivity by adding content during testing or having movable and interchangeable bits of the prototype. Additionally, using apps like Marvel[1] can create an actual app from sketches that can be send to testers. Digital wireframing tools like Pidoco[2] or Mockingbird[3] provide a variety of building blocks to build slightly more sophisticated wire framed screens and clickable prototypes often allowing a presentation as web-page or smart phone app.

For more sophisticated UI prototypes that provide a sense of the actual app design (e.g., for presentations to management or stakeholders) tools like Keynotopia[4] enable the team to build clickable UI prototypes in the actual design within Powerpoint or Keynote. Alternatively, HTML Pages can be created for the same purpose. The qualitative and measurement-based testing techniques presented in Sect. 4.2.2 are also usable during this phase.

When the InnoDev Team starts to look for aspects to clarify, to refine the user stories, or to decide on which features to include in the MVP, user story maps are a helpful tool. During user story mapping the functionality and features of the solution concept are transferred to agile user stories, which are then arranged on a User Story Map. Such a map arranges the main activities possible in the software from left to right in an order that makes sense, e.g. in a workflow or by priority (Patton 2009). Additionally, task centric user stories are arranged under the activity they belong to, also arranged from left to right. Tasks that can occur in parallel will be placed vertically under one another. Thus, such maps provide information about the planned functionality of the system under development and its iterations and can be used to identify holes and omissions in a backlog and plan releases that deliver value to user and business.

The Business Model Canvas and the Lean Canvas are valuable tools when it comes to creating, testing and adapting the business model for the software under development. The Business Model Canvas was originally proposed by Osterwalder (Osterwalder and Pigneur 2010) and has since been adapted for various special

[1] https://marvelapp.com/

[2] https://pidoco.com/

[3] https://gomockingbird.com

[4] https://keynotopia.com

cases. One such adaption is the Lean Canvas proposed by Maurya (2012), which is based on the work of Ries (2011).

Technical spikes are a good way to test software libraries and prototype the technology stack. Furthermore, they facilitate the possibility to work out solutions for technical issues or validate the technical feasibility of an idea through a simple implementation that is not aimed to be deliverable.

Once an MVP has usable features, it is a good idea to look for early evangelists (Blank 2007), that is users that are willing to take a risk and use an un-finished product. Such users provide crucial help in product development by their motivation to solve an urgent problem, encouraged by the vision of such a solution in place in the future.

In addition to the Scrum meetings as described in Sect. 4.1, a daily clickthrough of the current prototypes ensures that everyone in the team is up to date on the explored concepts and findings.

4.3.3 Roles

During this phase, the InnoDev Team is responsible for the planning and execution of the development sprints. Ideally the team that was working during the Design Thinking phase will continue during this phase and be extended with additional developers or other team members from areas of expertise as necessary for the software under development (e.g., back end developers, front end developers, database experts, UI developers, UI designers or interaction designers, business experts, or sales and marketing personnel).

(Potential) users and stakeholders will be responsible for testing the different prototypes developed during the Initial Development phase and give feedback on other artefacts and ideas, such as the business model.

Similarly, the project sponsor(s) give(s) feedback on the developed prototypes and the general direction of the project. In addition, they aim to facilitate communication with relevant departments of the their company and advertise the project progress to people interested.

During this phase, the role of the Product Owner (PO) starts to make sense. The PO is the representative of the customer and is responsible for creating the backlog and updating and prioritizing the user stories. For InnoDev, we propose to draw the PO or POs from the team members involved during the Design Thinking phase (e.g., a user researcher or designer trained for this role). A team of POs with a combination of designer, business and developer perspectives can be valuable for larger projects.

The Process Master has the same responsibilities as defined in Sect. 4.2.3.

4.3.4 Deliverables

The main deliverable of this phase is the MVP. It is complemented by other design and technical prototypes that are created for further refinement of the solution as well as a business model. The knowledge gained from developing and testing the prototypes, the business model and the MVP should lead to further functional, technical and non-functional requirements as well as more refined and new agile user stories.

4.4 Development Phase

The Development phase as illustrated in Fig. 6 is basically an agile development phase making use of the Scrum process framework and Lean Startup validation and scaling techniques. This phase enables the InnoDev Team to work towards a final product and business in incremental steps. Design Thinking is less prominent as in the other two phases. Instead of providing the main activities from its process, it only provides methods from its tool box where applicable.

4.4.1 Activities

The activities during this mode follow a basic software development approach using Scrum project management complemented by Lean Startup validation and scaling. The InnoDev Team focuses on the development of software increments including deployment and maintenance concepts as well as scaling the business. In case a) new features become necessary, b) existing features need to be refined or c) problems arise with existing features, the business model, or the team and their processes Design Thinking activities in the form of smaller workshops or single techniques can be chosen by the team to help them solve the task or problem at hand.

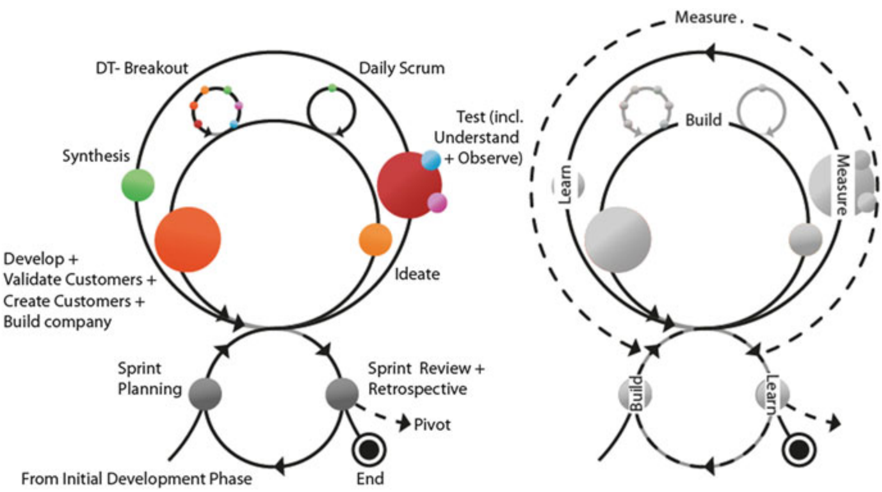


Fig. 6 Overview of the Development phase and representation of the build measure and learn concepts within this phase

4.4.2 Techniques

As this phase aims to further develop the MVP into a fully functioning software product and to develop the business around this product, software engineering techniques and customer development dominate the work. These include practices such as test-driven development, continuous integration, different review techniques to maintain code quality, collective code ownership, and continuous customer testing [compare (Beck 2000; Ries 2011)]. Additionally, techniques from Lean Startup used to scale the business and for validation are helpful. These include making use of actionable metrics, for acquisition, activation, retention, revenue and referral (AARRR), which can be used to assess the product performance and evaluate the product, the business model and the marketing strategy (Maurya 2012). As this phase is still about continuous learning and improvements, a good technique to find out which of two implementations works better A/B or split testing is helpful. It can be used to evaluate different marketing campaigns in different areas or different features or implementations of a feature by splitting the users into groups and providing them with different versions. After acquiring customers, and looking ahead for growth it would be a good move to establish a customer advisory board.

Design Thinking techniques will be used in an ad-hoc manner as necessary and therefore can be drawn from the entire spectrum of techniques, for example the techniques described in Sect. 4.2.2.

4.4.3 Roles

The responsibilities of the InnoDev Team during this mode are similar to those in the preceding mode. They plan and execute the sprints implementing functional software increments. If needed, additional Scrum teams can be added to allow for parallel development.

The (potential) users are again tasked with testing the software increments and giving feedback to changes and feature ideas.

The Project Sponsor is still tasked with facilitating communication with interested departments inside their company and to promote the project progress to interested parties. Additionally, he will give feedback on the developed software increments during reviews.

The Product Owner has similar responsibilities as described in Sect. 4.3.3.

The Process Master has the same responsibilities as during the other modes.

4.4.4 Deliverables

The Development phase focuses on creating tested, working software and developing the corresponding business. Thus, the deliverables are a product that is continuously improving as well as a growing business.

During this phase, all software development should be potentially shippable by the company. This means that the software should adhere to product standards as defined by the company and necessary for the market and the users. Furthermore, it needs to be delivered to the users or customers on a regular basis even if it is not an online product or a mobile app, in which cases a deployment strategy might be necessary.

4.5 Configurations

InnoDev is a software development methodology generally applicable to different company settings such as startups, small and medium-sized enterprises and large organizations. However, differences in a specific context, such as team-size, project goals, product size, level of expertise for the methodology etc. exist and should be targeted by adapting InnoDev to the specific context of a project. In this section, we will describe possible adaptations to the general InnoDev process for specific needs which we believe should be addressed. Our suggestions in this chapter do not present a complete list of possible adaptations but rather stem from our former work and can be extended for other needs and contexts.

4.5.1 Goals

We believe our process to be applicable for the development of innovative and new software products as well as for the incremental and on-going development of existing software products.

When a new product is developed, the challenge used to start into the Design Thinking phase can be formulated accordingly, giving the general area of the product and some context of its users, an example challenge could be: “How might we help elderly people to be more mobile in their everyday life?”. The team should then start to investigate the market for existing products in this area and research user needs by interviewing and observing the user group.

In case existing software will be extended and InnoDev is used to discover new features, the InnoDev Team needs access to the existing software or, ideally, to include members from the development team of the existing software. In such cases, company requirements or product standards may already play a larger role in the beginning of the InnoDev process.

The challenge (or problem statement) should be formulated accordingly, for example: “Discover features for our online-shop that appeal to teenagers!”. The InnoDev Team then starts with the same process but is already more focused on the existing solution and might have access to existing customers for interviews and observations. Furthermore, the development of a business model during the Initial Development phase might not be necessary if the existing business model is sufficient.

4.5.2 Team Size and Setup

The team size has a potential impact on the project. Small teams can be more flexible and more easily able to change their direction, whereas large teams have a greater work force and thus the ability to cover more aspects or produce more results. In the case of working with a bigger team comprised of several smaller teams, it is necessary to scale the InnoDev process, establish a communication structure between teams and some form of a team lead. For ideas on how to scale InnoDev please refer to Sect. 4.6.

The team setup could also be configurable. Ideally, the InnoDev Team stays constant throughout the process to avoid handovers and the not-invented-here syndrome. This state can be achieved by keeping everyone who was involved during the Design Thinking phase in the team for the later phases and only adding personnel as needed, for example if special skills or more development power are necessary. However, that might not always be possible due to people leaving the companies or having other priorities inside the company. Switching teams between phases is therefore possible, however in such a case special care needs to be taken of producing light-weight documentation accompanying each deliverable to make sure the new team understands where ideas are coming from, why features are important, and so on.

4.5.3 Level of Expertise

The InnoDev process can be used by teams that have all the required expertise to use Scrum, Lean Startup, and Design Thinking as well as inexperienced teams who only have some or none of that expertise at their disposal. Experienced teams will be able to choose the right techniques during each activity and decide when to move from one activity to another or when to switch to the next phase by themselves or with the help of the InnoDev Process Master. The decisions to switch between phases can be made in consultation with project leads or managers, should they be established. More inexperienced teams will probably need stronger guidance in making these decisions. In such cases, the Design Thinking phase could prescribe techniques that the team has to run through, as is proposed by de Paula (2015). Another possibility is to structure both the Design Thinking and the Initial Development phase with the help of milestone deliverables as proposed by Vetterli et al. (2013). Figure 7 visualizes a possible distribution of such milestones in the Design Thinking phase and the Initial Development phase. Finally, inexperienced teams can be guided and supported by a team of coaches and experts throughout the process if the necessary budget and manpower are available.

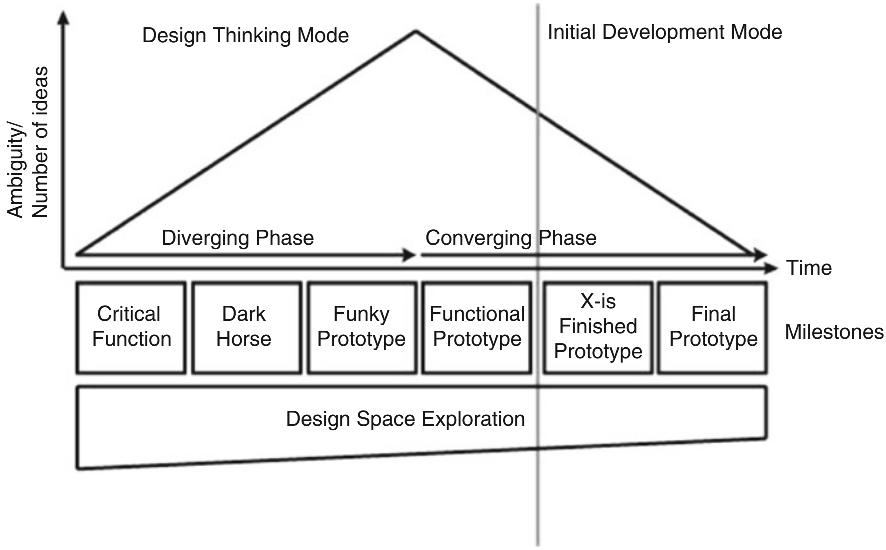


Fig. 7 Milestone concept during the Design Thinking and the Initial Development phase (Häger et al. 2015 adapted from Vetterli et al. 2012)

4.5.4 Company Sizes

The size and structure of a company can have a large impact on a project. While teams in smaller companies can work relatively free and in close cooperation with management. Teams in large organizations might have to report to various managers, fit into the company strategy, adhere to quality and security standards and be subject to audits. In such cases, the Initial Development phase should be used to create specifications of how the product under development will integrate into the company context, including the identification of dependencies with other projects, or possibilities to reuse existing software systems or components in the final implementation. Furthermore, it is possible to set up a transition between the phases of InnoDev in a stage-gate manner, thus allowing for management reporting and approval, as well as audits, before moving to the next phase.

Another smaller aspect that might depend on the company size, is the fit of specific techniques for the teams. For example, the development of a business model is a crucial aspect no matter the company's size, but different tools are implemented depending on whether a company is large or small. Based on experiences from practice the Lean Startup canvas is useful for startups, while the business model canvas is the right tool for larger organizations.

4.6 Scaling InnoDev

If the software under development is large in terms of features and potential user groups (e.g., business software that aims to bundle several business processes for several users into one piece of software), working with one small team might not be sufficient. In such cases, it becomes necessary to scale InnoDev to be able to work with several teams.

Scaling for the Design Thinking phase is scarcely researched so far. One possible way to scale the Design Thinking process has been reported by Häger and Teusner (2014). They describe a multiteam Design Thinking workshop series to kickstart larger software projects. Key elements of their approach are depicted in Fig. 8.

A kickoff workshop introduces the teams to their challenge and if necessary to Design Thinking. During this workshop, all teams fast forward once through the whole Design Thinking process. Following the workshop, the teams have time for teamwork intertwined with further workshops, in which they run through the Design Thinking process again with more time and iterate on their ideas. The workshops allow for an exchange of ideas and feedback as well as communication between the different teams. Furthermore, they provide a possibility to track the progress of the teams. Once the teams have arrived at final ideas and prototypes, Häger and Teusner propose to evaluate ideas and combine them to possible larger pieces of software that can then be further evaluated and developed in follow-up projects. Figure 9 depicts a possible development from ideas to smaller projects and then to a final bigger project and maps those projects to the phases of InnoDev.

Smaller follow-up projects include innovation projects, technical proofs of concept, or developing MVPs for specific user groups and help to clear open technical or conceptual questions and further refine specific ideas. Finally, ideas and outcomes of these smaller follow-up projects will be combined into one bigger software development project.

We believe the approach presented by Häger and Teusner can be integrated into the Design Thinking phase, as the prerequisites, activities, deliverables and roles that are necessary largely overlap. Only the role of the team lead or team leaders would be new and the workshops would form another activity in addition to the activities for each of the participating InnoDev Teams. We believe that the team leaders should be enlisted out of the InnoDev Teams similar to a group of POs in a scaled Scrum environment, thus allowing all teams to be represented. The concept of the follow-up projects also fits well into the Initial Development phase making it possible to scale this phase in the form of multiple projects that refine and test different parts of the final combined products. Ideally the InnoDev Teams from the first phase will be able to continue working on their ideas during the Initial Development phase.

Finally, when moving towards the bigger software development project, teams working in the Development phase can rely on methodologies for Scaling agile software development, such as Large-Scale Scrum (LeSS) (Larman and Vodde 2009, 2010, 2013) or the Agile Scaling Model (Ambler 2009). Both Ambler, and Larman and Vodde (Larman and Vodde 2009, 2010, 2013) present examples and

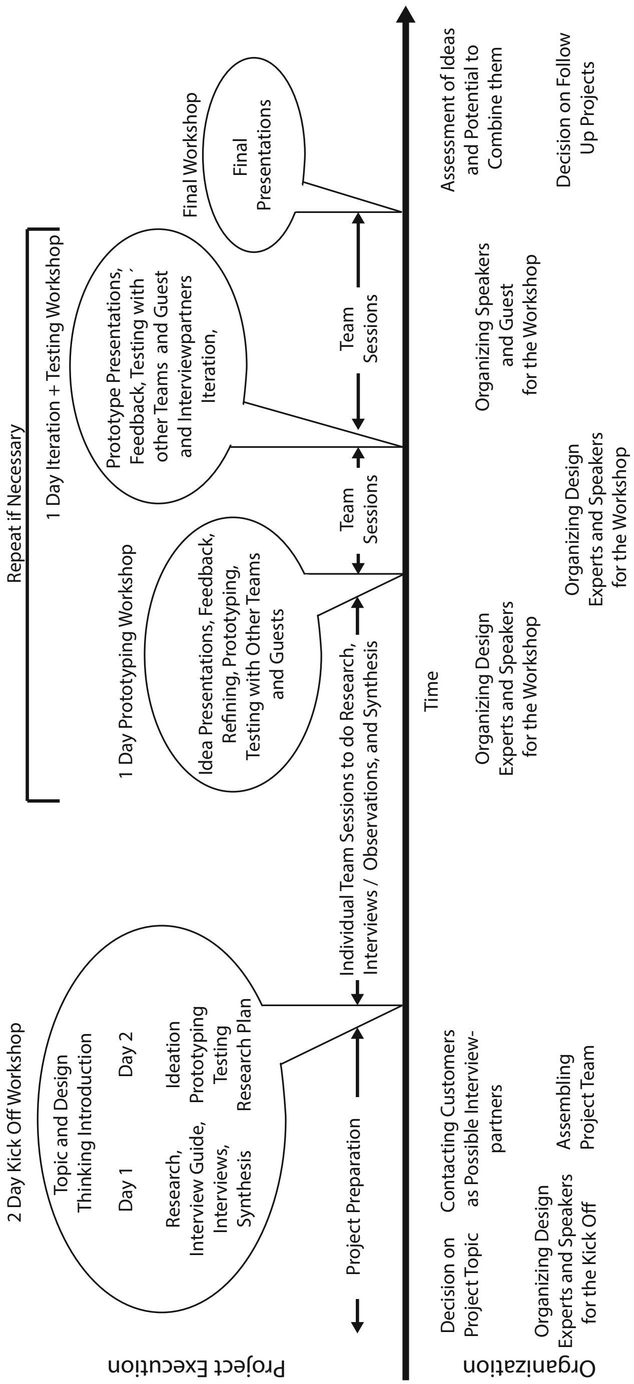


Fig. 8 Scaling Design Thinking for several teams in a series of workshops (Häger and Teusner 2014)

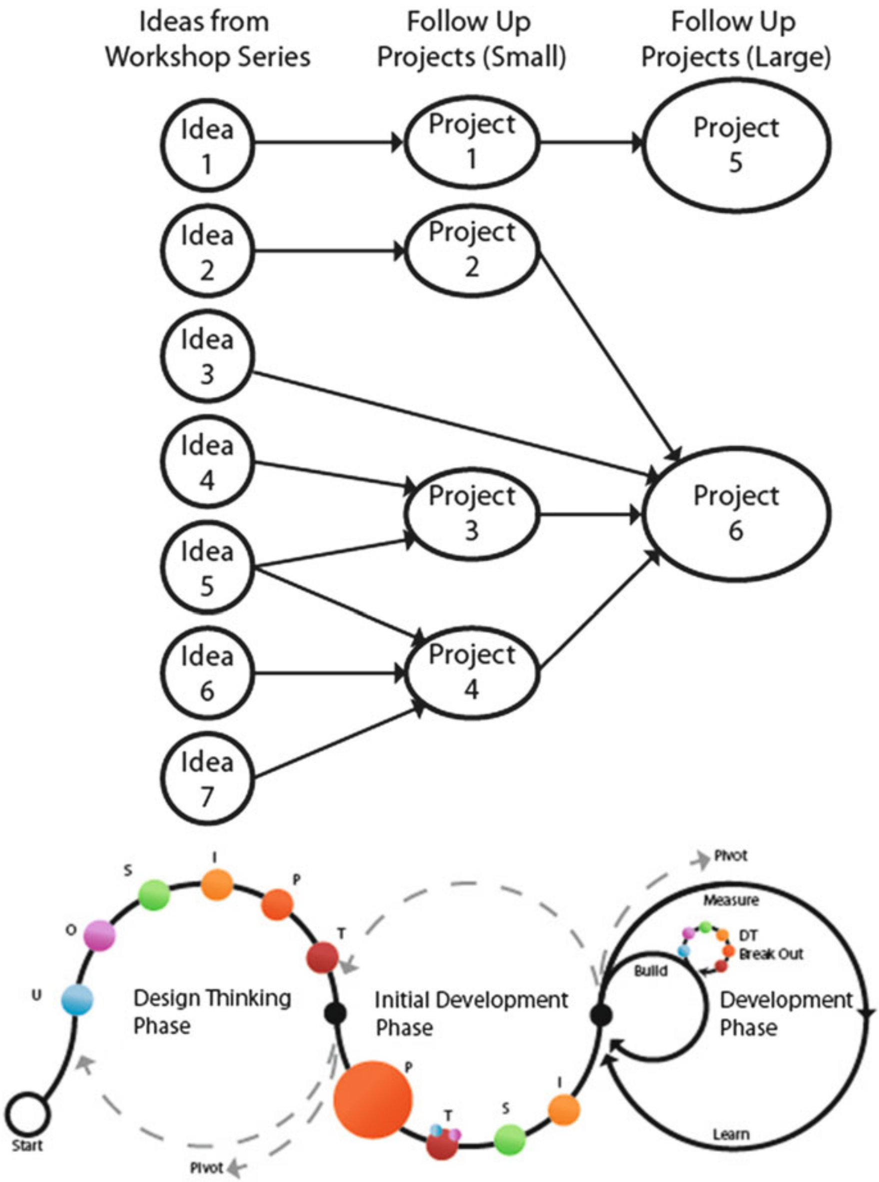


Fig. 9 Development of ideas from the workshop into a series of follow-up projects mapped to the InnoDev phases (adapted from Häger and Teusner 2014)

case studies on how agile processes can be scaled for large project teams and explain appropriate techniques. Example techniques include the Scrum of Scrums or Meta Scrum, communities of practice or learning days.

In the Scrum of Scrum technique, the individual Daily Scrum of all teams is followed by a Daily Scrum of Scrums with an ambassador from each team, who will give a progress report from his team and take back important information to his team members. If necessary, this technique can be used on multiple levels. Communities of practice allow for the exchange of knowledge and ideas or the discussion of problems between people with the same role or type of expertise. In this technique, such groups meet regularly to discuss problems and ideas with each other. Example groups could be design (incl. UI, interaction and visual designers), testing (including various testers), DevOps, or business (incl. management, sales and marketing personnel). Learning days or learning workshops are a way to spread knowledge throughout big teams or companies, in which a team member or a team can share useful techniques, case studies, or other interesting knowledge with other teams.

5 Evaluation

It has been demonstrated that integrating Design Thinking to the software development process enables increased team collaboration (Carlgren et al. 2014), better understanding of the user and product (Liedtka 2015). For this study, we aim to evaluate InnoDev by following a two-phase approach: survey development, and survey application and workshops. Survey Development corresponds to the creation and measurement of an instrument to validate the InnoDev model. Survey application and workshops aim to run workshops that teach InnoDev with different companies and apply the validated instrument to developers, designers and managers from the companies attending the workshop. Additionally the questionnaire will be send out to various companies involved in software development to collect a large sample of responses.

5.1 Survey Development

In order to validate InnoDev, a questionnaire will be developed, validated, tested and applied to developers, designers and managers in software companies. The items in the questionnaire will be developed to answer the following research questions:

RQ1: To what extend are solutions proposed by InnoDev important to software development companies?

RQ2: To what extend are solutions proposed by InnoDev already implemented by software development companies?

RQ3: To what extend can solutions proposed by InnoDev help avoid common flaws in the software development process?

In order to ascertain that our questionnaire is well designed and the items are measuring all aspects of the InnoDev model we will make use of content validity and face validity checks. A content validity check will be undertaken to ascertain whether the content of the questionnaire is appropriate and relevant to the study purpose. Six experts from the areas of Design Thinking, software development, and survey design will be asked to review the items of our questionnaire to ensure they are consistent with InnoDev. A face validity check verifies whether the questionnaire is appropriate to the purpose of the study and content area. It evaluates the appearance of the questionnaire in terms of feasibility, readability, consistency of style and formatting, and the clarity of the language (Haladyna 2004).

5.2 Survey application and workshops

In order to apply the survey, we will distribute the questionnaire to a wide range of software development companies through appropriate groups on LinkedIn and Xing and through personal contacts. Additionally, we will design a workshop aimed at walking companies through each step of the InnoDev model in order to teach the InnoDev process and to apply our questionnaire to the attendees. The workshop will be designed to accommodate the needs of a wide variety of software company roles. For example, designers, developers, managers and stakeholders who have a keen interest in software solution innovation. Participants in the workshop will benefit from the applied learning of new techniques through the InnoDev model. The workshop will encourage participants to identify (a) opportunities for new software solutions in their companies and / or (b) enhance existing software and service offerings by aligning solutions to specific customer needs. Specifically the workshop will demonstrate how InnoDev can support companies to align Design Thinking, product development, customer development and value realisation for new software products and services.

Before running the workshop we will conduct a pilot test of the workshop. This pilo test will be a condensed version of the final workshop. The participants of the pilot workshop will be asked to answer a questionnaire after the pilot workshop to evaluate it's quality. The questionnaire will include a simplified set of criteria for evaluation as used by (Pigosso et al. 2013), which are: utility, consistency, simplicity, clarity, coherence, instrumentability and forecast. Based on the results of the pilot workshop the concept will be refined. We plan to run at least two workshops with startup companies from Galway, Ireland and Berlin, Germany.

6 Outlook and Summary

The aim of this chapter has been to describe in detail a framework that combines Design Thinking, Lean Startup and Scrum for software development that can deliver the innovative customer-oriented products and services required by competitive companies. InnoDev was developed based on existing models from the literature that have been recently studied. The findings from our study are a step towards aligning relevant research in order to enable the next generation of research on the software development process.

First, we describe the three phases of InnoDev and how Design Thinking, Lean Startup and Scrum interact with each other: The Design Thinking phase, the Initial Development phase and the Development phase. All three phases are in line with what others researchers (Grossman-Kahn and Rosensweig 2012; Hildenbrand and Meyer 2012) claimed to be relevant to a software development process. The three phases essentially center on trying and learning each using different tools and techniques. Design Thinking aims to understand and learn about problems and user needs in order to derive solutions and product options. Lean Startup aims to learn about business strategies and scaling options and Scrum about development options and further directions during software development with changing requirements.

Further, we propose that Scrum is used as the overall project management method underlying all phases of InnoDev. In particular, we propose companies use Scrum to structure the Design Thinking phase in order to let teams get a feeling for the duration and value of Design Thinking activities, and to enable them to better structure their creative work.

Our findings provide complementary perspectives regarding software development strategies, roles and techniques. Future work could expand our findings and evaluate InnoDev in an industry scenario, which might help us better understand how to enhance the synergy between the approaches.

This study advances knowledge of Design Thinking and software development by providing a detailed description of a tool that combines best practices for creating more innovative software products. The results of this investigation can help managers to evaluate their software development process and thereby improve its effectiveness and create more efficient user-driven solutions.

References

Ambler, S. W. (2009). The agile scaling model (ASM): Adapting agile methods for complex environments. *Environments*.

Beck, K. (2000). *Extreme programming explained: Embrace change*. Boston: Addison-Wesley Professional.

Beverland, M., & Farrelly, F. J. (2007). What does it mean to be design-led? *Design Management Review, 18*(4), 10–17.

Blank, S. G. (2007). *The four steps to the epiphany: Successful strategies for products that win*. California: S. G. Blank.

Carlgren, L., et al. (2014). Exploring the use of design thinking in large organizations: Towards a research agenda. *Swedish Design Research Journal, 1*(14), 47–56.

Chin, K.-S., et al. (2008). Group-based ER–AHP system for product project screening. *Expert Systems with Applications, 35*(4), 1909–1929.

d.school Stanford. Bootcamp Bootleg.

de Paula, D. F. O. (2015). Model for the Innovation Teaching (MoIT): um modelo baseado em Design Thinking, Lean Startup e Ágil para estudantes de graduação em computação.

de Paula, D. F. O., & Araújo, C. C. (2016). *Pet empires: Combining design thinking, lean startup and agile to learn from failure and develop a successful game in an undergraduate environment*. In International Conference on Human-Computer Interaction (pp. 30–34). Springer.

de Paula, D. F. O., et al. (2014). *Building a quality mobile application: A user-centered study focusing on design thinking, user experience and usability*. In: International Conference of Design, User Experience, and Usability (pp. 313–322). Springer.

Deemer, P., Benefield, G., Larman, C., & Vodde, B. (2012). The Scrum Primer: A lightweight guide to the theory and practice of scrum. Version 2.0.

Derby, E., et al. (2006). *Agile retrospectives: Making good teams great*.

Dingsøyr, T., et al. (2012). A decade of agile methodologies: Towards explaining agile software development. *Journal of Systems and Software, 85*(6), 1213–1221.

Dobrigkeit, F., & de Paula, D. F. O. (2017). *The best of three worlds – The creation of InnoDev a software development approach that integrates design thinking, scrum and lean startup*. In Proceedings of the 21st International Conference on Engineering Design (ICED17), Vol. 8, Human Behaviour in Design (pp. 319–328). Vancouver, Canada.

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology, 50*(9–10), 833–859.

Eisenmann, T. R., et al. (2012). Hypothesis-driven entrepreneurship: The lean startup. *Harvard Business School Background*. Note 812-095.

Erickson, J., et al. (2005). Agile modeling, agile software development, and extreme programming: The state of research. *Journal of Database Management, 16*(4), 88.

Freeman, R. E. (2010). *Strategic management: A stakeholder approach*. Cambridge: Cambridge University Press.

Frishammar, J., et al. (2011). Beyond managing uncertainty: Insights from studying equivocality in the fuzzy front end of product and process innovation projects. *IEEE Transactions on Engineering Management, 58*(3), 551–563.

Grenning, J. (2002). Planning poker or how to avoid analysis paralysis while release planning. *Hawthorn Woods Renaissance Software Consulting, 3*, 1–3.

Grossman-Kahn, B., Rosensweig, R. (2012). Skip the silver bullet: Driving innovation through small bets and diverse practices. *Leading Through Design*, 815–829.

Gutzwiller, T. A. (2013). *Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen*. Berlin: Springer.

Häger, F., & Teusner, R. (2014). *Scaling design thinking – Using a multi-team design thinking workshop to kick start software projects*. Presented at the Design Thinking in Business Information Systems Workshop at the European Conference on Information Systems, Tel Aviv.

Häger, F., et al. (2015). DT@Scrum: Integrating design thinking with software development processes. In H. Plattner et al. (Eds.), *Design thinking research* (pp. 263–289). Heidelberg: Springer.

Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hildenbrand, T., & Meyer, J. (2012). Intertwining lean and design thinking: Software product development from empathy to shipment. In A. Maedche et al. (Eds.), *Software for people: Fundamentals, trends and best practices* (pp. 217–237). Heidelberg: Springer.

Hirschfeld, R., et al. (2011). Agile software development in virtual collaboration environments. In *Design thinking* (pp. 197–218). Heidelberg: Springer.

Kerth, N. L. (2000). *Project retrospectives – A handbook for team reviews*. New York: Dorset House.

Komus, A. (2017). Status quo agile. Stud. Zur Verbreit. Nutzen Agil. Methoden Hochsch. Kobl.

Kua, P. (2013). *The retrospective handbook*. E–book Available Httpsleanpub Com–retrospective–handbook.

Larman, C., & Vodde, B. (2009). *Scaling lean & agile development*.

Larman, C., & Vodde, B. (2010). *Practices for scaling lean & agile development: Large, multisite, and offshore product development with large-scale Scrum*. Upper Saddle River, NJ: Addison-Wesley.

Larman, C., & Vodde, B. (2013). Scaling agile development. *CrossTalk, 9*, 8–12.

Liedtka, J. (2015). Perspective: Linking design thinking with innovation outcomes through cognitive bias reduction. *Journal of Product Innovation Management, 32*(6), 925–938.

Lindberg, T., et al. (2011). Design thinking: A fruitful concept for IT development? In C. Meinel et al. (Eds.), *Design thinking* (pp. 3–18). Heidelberg: Springer.

Maurya, A. (2012). *Running lean: Iterate from plan A to a plan that works*. Sebastopol: O'Reilly Media.

Müller, R. M., & Thoring, K. (2012). *Design thinking vs. Lean startup: A comparison of two user-driven innovation strategies*. In Leading innovation through design proceedings of the DMI 2012 International Research Conference 2012, 8–9 August, Boston. Boston: Design Management Institute.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers*. Hoboken: Wiley.

Patton, J. (2009). Telling better user stories – Mapping the path to success. *Better Software, 11*(7), 24.

Pigosso, D. C. A., et al. (2013). Ecodesign maturity model: A management framework to support ecodesign implementation into manufacturing companies. *Journal of Cleaner Production, 59*, 160–173.

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. New York: Crown Business.

Schwaber, K. (1997). Scrum development process. In *Business object design and implementation* (pp. 117–134). London: Springer.

Schwaber, K., & Beedle, M. (2001). *Agile software development with scrum*.

Schwaber, K., & Sutherland, J. (2013). *The scrum guide – The definitive guide to scrum: The rules of the game*.

Scrum Alliance. (2016). *The 2016 state of scrum report*. Scrum Alliance.

Thoring, K., & Müller, R. M. (2011). *Understanding design thinking: A process model based on method engineering*. In DS 69: Proceedings of E&PDE 2011, the 13th International Conference on Engineering and Product Design Education, London, UK, 08.-09.09.2011.

Verganti, R. (1997). Leveraging on systemic learning to manage the early phases of product innovation projects. *RD Management, 27*(4), 377–392.

Version One. (2017). 11th annual state of agile report. Technical report, Version One.

Vetterli, C., et al. (2012). Initialzündung durch Embedded Design Thinking – Ein Fallbeispiel aus der Finanzindustrie und wie dadurch ein Wandel in der Innnovationskultur einer IT-Abteilung eingeleitet wurde. *OrganisationsEntwicklung: Zeitschrift für Unternehmensentwicklung und Change Management, 31*(2), 22–31.

Vetterli, C., et al. (2013). *Jumpstarting scrum with design thinking*.

Vilkki, K. (2010). When agile is not enough. In *Lean enterprise software and systems* (pp. 44–47). Heidelberg: Springer.

Wölbling, A., et al. (2012). Design thinking: An innovative concept for developing user-centered software. In *Software for people: Fundamentals, trends and best practices* (pp. 121–136). Berlin: Springer.

Towards Exploratory Software Design Environments for the Multi-Disciplinary Team

Patrick Rein, Marcel Taeumel, and Robert Hirschfeld

Abstract The creation of a new software system can be a wicked problem. Consequently, it is important for such projects to have a collaborating team of experts from multiple disciplines. While agile development processes foster such a collaboration on the social level, the tools used by individual experts still prevent team members from seeing the overall result of their collective modifications on the resulting system. Roles in the process, such as content designers and user experience designers, only get feedback on the impact of their changes on their artifacts. Based on the concept of exploratory programming environments, we propose a new perspective on the environments used in software development, called *exploratory software design environments*. We describe the properties of such an environment and illustrate the perspective with existing related tools and environments.

1 Introduction

Software development has the properties of a wicked problem: Requirements might only become apparent after an interim solution was proposed and software is "never done" as the intended real-world use cases for the software constantly change (Rittel and Webber 1973; Conklin 2006; DeGrace and Stahl 1990). Further, the creators of a software system have to account for a variety of properties such as technical stability and maintainability, usability of the user interface, correctness of the domain model, and actual usefulness for the users. Consequently, software development can benefit from insights of the Design Thinking methodology, in particular the consideration of multiple viewpoints for solving such wicked problems (DeGrace and Stahl 1990; Beck 2000).

In order to create an appropriate solution, a multi-disciplinary team has to closely collaborate, as only then can the multiple perspectives of the participants actually

P. Rein (✉) · M. Taeumel · R. Hirschfeld
Software Architecture Group, Hasso Plattner Institute for Digital Engineering, Potsdam, Germany
e-mail: patrick.rein@hpi.de; marcel.taeumel@hpi.de; robert.hirschfeld@hpi.de

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_12

contribute to the design. In such close collaboration, team members are not only interested in finishing their individual tasks but *continuously assess the impact of their own contributions* on the overall design and comments on any other contribution if they think it necessary. A commonly described factor for creating such a team culture is the creation of a common purpose within the team.

Software development can benefit from teams of experts that incorporate multidisciplinary knowledge. The variety of properties of a software system makes it essential for the design process that a variety of people with different backgrounds are involved in the creation of the software, such as back-end developers, user interface designers, actual users, and experts of the application domain. The common purpose of such software development teams is ideally the collective and continuous evolution of a system that brings value to its users. Agile processes are based on the notion of a team sharing a common purpose. These processes try to support the team culture through appropriate techniques. For example, Extreme Programming (XP) lists "The Whole XP Team" as one of its practices and describes it as if ". . . they [the team] were roped together. Walking abreast, they could make more progress than if any one group tried to force the others to follow." (Beck 2000).

While such development practices aim to support a culture of working together on a single system, this culture is often not reflected in the software tools used by individual participants. All experts operate their own tools, creating an output which is only later combined into the running system. For example, technical writers are often passed a file with a long list of placeholders which they should, for example, translate into full-length labels. The effects of changes to the text might only first become visible to the other team members much later in the process and only in case where they actually run the new system version. Consequently, even in cases where an agile development process might aim at collaboration on one system, the tools only allow for cooperation almost resembling a *software factory* with single workstations (see Fig. 1).

Instead, we should aim for a *software workshop* in which all participants of the process work together on the actual, running system (see Fig. 2). In such a workshop, a technical writer would change the labels while the software is running and others would shortly afterwards see the changed labels as well. Whenever someone applies a change to the system, the effect should be visible to the other team members shortly thereafter. This workflow facilitates a sense of working together on one system and makes collaboration more likely.

In this article, we illustrate the factors impeding collaboration during the design of software in traditional environments. Further, we show how so-called *exploratory programming environments* can serve as a foundation for a software workshop in which people with different roles can collaborate directly on one system. We do so by describing the properties of exploratory programming environments and illustrate these properties with two exemplary programming systems. We then generalize these properties as properties for *exploratory software design environments,* which provide an exploratory workflow for all participants of the process. To show how such an environment might work we further describe a number of exemplary tools and environments that implement characteristic aspects.

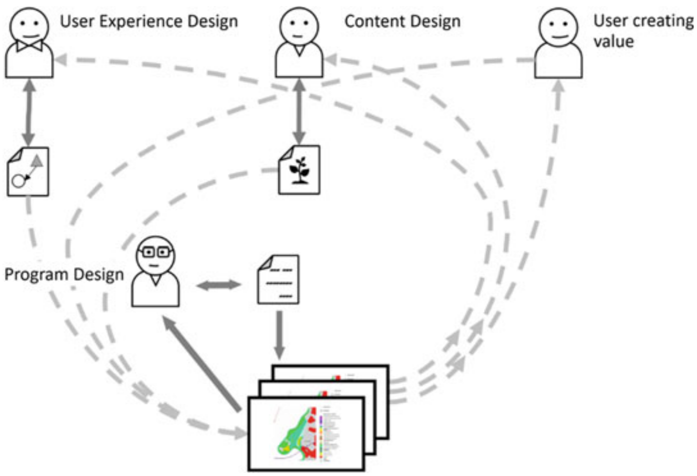


Fig. 1 The current workflow in software development. Most team members work on just their type of artifacts (bold arrows). They only get feedback on the impact of their changes on the system at a later point in time—first after submitting their artifacts to the program designers (dashed lines)

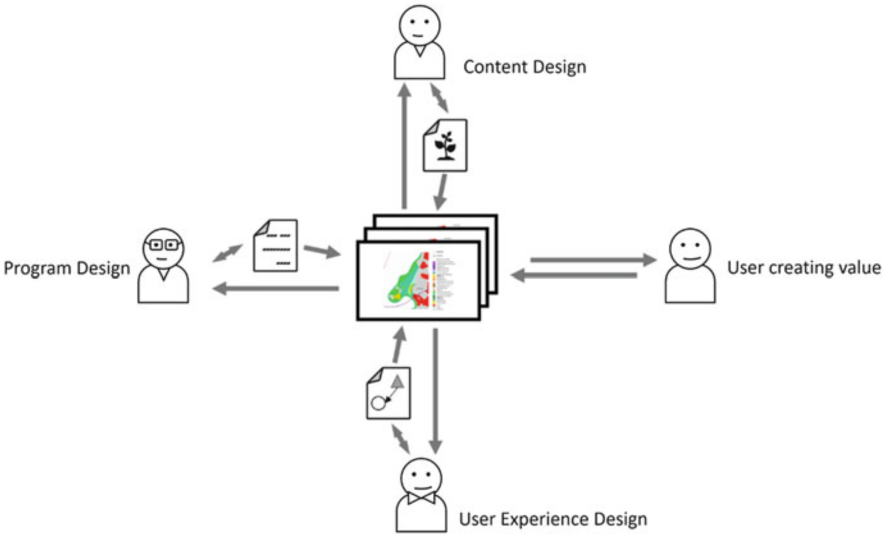


Fig. 2 The ideal workflow in exploratory software design environments for the whole team. Every team member can work on their artifacts but also get direct feedback on their modifications as they relate to the system as a whole. Further, they can see how their modification interacts with modifications of others

2 From Cooperation to Collaboration in Software Development

Agile processes such as Extreme Programming (XP) share a number of principles and values with the design thinking methodology. Both are iterative in nature, make creating value for the user a primary goal, and emphasize self-sufficient and multi-disciplinary teams. To actually leverage the different viewpoints of a multi-disciplinary team, team members have to collaborate beyond mere cooperation. As software development already entails particular tasks, a multi-disciplinary software development team has a set of artifacts and tools that correspond to typical roles: content creation, user experience design, program design, and user.

2.1 Design Thinking, Wicked Problems, and Agile Processes

Wicked problems are problems with "no definitive formulation" due to requirements which are incomplete from the start or might change during the design of a solution. Traditional examples of wicked problems are social problems such as drug abuse and homelessness. In a more general way, they are defined by a set of six characteristics (Conklin 2006) [derived from a larger catalogue of 11 characteristics (Rittel and Webber 1973)]:

1. The problem is not understood until after the formulation of a solution.
2. Wicked problems have no stopping rule.
3. Solutions to wicked problems are not right or wrong.
4. Every wicked problem is essentially novel and unique.
5. Every solution to a wicked problem is a 'one shot operation.'
6. Wicked problems have no given alternative solutions.

The design thinking process is suitable for approaching wicked problems (Buchanan 1992). For example, the iterations in the design process allow team members to refine their understanding of the problem after each iteration (characteristic 1). Further, the novel and unique nature of the problem is covered by techniques for ideation to support the team in creating new and fitting solutions (characteristic 4). Even early design thinking methodologies already focused on similar types of problems (Arnold 1956, 1959/2016). Further, wicked problems consist of interdependencies of various individual factors. Each factor might only be understood in terms of a particular field, such as sociology, art, and mechanical engineering. Thus, design teams should ideally consist of experts from a variety of fields. Due to their different backgrounds, each team member has a different perspective on the original problem and can hence determine sub-problems related to their field.

The design of software systems is also considered a wicked problem (DeGrace and Stahl 1990). A major aspect which makes software development wicked is that the actual requirements for a software system are only understood after parts of the software have been built and are actually in use. Often, users require an intermediate state of the software to determine what they actually need. Further, software also does not have a "stopping rule" (characteristic 2). As the context of its use is constantly changing, software has to adapt accordingly. This is summarized by the saying in software industry that "software is never done".

Agile processes acknowledge the wicked nature of software development. One of the principles of Extreme Programming (XP) is, for example, "embrace change," meaning that development should happen in small iterations to get feedback from users in a timely manner. Based on this feedback, the system can then directly be adapted to best fit the new requirements (Beck 2000). In this respect, every iteration of an XP team is a design iteration. At the beginning, needs and wishes from the users are collected. The team then works out a solution for these challenges and produces a small increment in the features of the system. This increment serves as a prototype which is directly evaluated with the users by incorporating it into the running software. Observations and feedback from users are directly incorporated into the next iteration. Also, software development covers more than the mere production of source code. Software development also covers activities such as user experience design, interface design, and content creation by domain experts. Consequently, XP also emphasizes a "whole team" and collaboration becomes paramount in the process. Every activity or aspect of the software should be covered by someone in the multi-disciplinary team.

2.2 From Cooperation to Collaboration

For multi-disciplinary teams to be effective, they have to *collaborate* on solving the problem and not only *cooperate*. Although both of the words "cooperation" and "collaboration" generally refer to working together, the style of working together they describe differs.

Cooperation is defined by Merriam Webster as "to act or work with another or others; act together or in compliance." The emphasis here is on team members merely *acting* together. This does not imply a shared goal or active support for fellow members. Individuals who cooperate have some overlap of their goals but the individual goal dominates. An example for cooperation are bureaucratic organizations. The group of people who make up the organization cooperate by each member fulfilling an individual task and thereby together providing the service of the organization.

Collaboration is a special form of cooperation defined by Merriam Webster as "to work jointly with others or together especially in an intellectual endeavor." People who collaborate closely work with others to achieve a common goal. Design teams typically collaborate as they discuss ideas in the group together. For a team to collaborate closely, it is necessary that each member has a shared understanding of the common purpose of the team. Further, every team member has to assess the complete situation continuously and put it in relation to the strategic goal of the team, similar to the way each soccer player has to continuously monitor the complete soccer field and not only concentrate on their "patch of grass." (McChrystal 2015).

In his book "Team of Teams", Stanley McChrystal does not mention the two terms cooperation and collaboration explicitly. Instead he illustrates the two styles of working together through a comparison of traditional command structures in the military and team-based structures in a variety of domains (McChrystal 2015). The author describes command structures as a way of cooperation: "[. . .] in a command, the leader breaks endeavors down into separate tasks and hands them out. The recipients of instructions do not need to know their counterparts, they only need to listen to their boss. In a command, the connections that matter are vertical ties." The author states that this cooperation in command structures is efficient, but at the same time rigid, which accordingly is not a good fit for modern challenges. The context and requirements of modern challenges change too quickly for any pre-determined plan to be applicable. Instead, organizations should focus on small empowered teams whose members collaborate on working towards a common goal. Again, one of the characteristics of such teams is that "team members tackling complex environments must all grasp the team's situation and overarching purpose [. . .] They must be collectively responsible for the team's success and understand everything that responsibility entails."

The design of a software system depends on such collaboration between experts from multiple disciplines. However, the fact that these team members are experts of their own discipline and masters of their own tools makes it easy for them to stick to just their own "patch of grass."

2.3 *The Whole Team: Multiple Disciplines for Multiple Perspectives*

Software development benefits from a pre-defined type of artifact to be produced: the software system. There are particular roles relevant to the software design process (Beck 2000), such as testers, interface designers, programmers, technical writers, and managers. Depending on the type of system to be created, different roles might be more active in the process than others. For example, in the development of a computer game, artists can make up more than half of the team. Similarly, when working on a software tool for a particular domain, domain experts might outnumber programmers (for example, the biochemistry software tool company Synthace lists two biochemistry scientists as technical leads and only one software engineer[1]). Enabling all the roles to participate equally during software design is beneficial for the quality of working together.

[1] https://web.archive.org/web/20171205131307/https://synthace.com/who-we-are/ accessed on 5th of December 2017.

Further, these roles are only approximate groupings of activities. For example, in XP "roles on a mature XP team aren't fixed and rigid. The goal is to have everyone contribute the best he has to offer to the team's success." (Beck 2000). We summarize these roles into the four categories:

- User
- User experience designer
- Content designer
- Program designer

We want to illustrate the typical contributions, tools, and artifacts each role creates in an iteration during the development process.

Users provide new requirements and general use cases for the software. They can have a precise idea of these qualities or they might only be able to visualize them roughly. Generally, users are the ones who generate value with the software system produced. They are affected by all the decisions of the other roles and at the same time provide the requirements, use cases, and the overall purpose of the software to be produced. Besides the direct users, we also include customers and domain experts in this role. In consumer software, users often interact with the system design team through issue tracking software. In more specialized software they might be able to talk in person to the design team and may even be able to join them during an iteration.

Early in the process *user experience designers* might first produce paper-based sketches of user interfaces to check with users whether this is what they need. Additionally, they can create storyboards to document a workflow users want to have supported in the system. In general, the user experience designers determine the actual interactions and feedback mechanisms of the system to make the program useful to the users. They take care of the intricacies of single user interfaces as well as the efficiency complete workflows throughout the system. Activities of this role might also be subsumed under the terms "usability engineer" or "user interface engineer." They work with a variety of tools, starting with pencil and paper. For visual design, user experience designers might use graphical editors in which they create screenshots of the future interface. They might also use user interface builder tools in which they can already define the actual user interface in a graphical manner.

Content designers generally create the texts, graphics, or pre-loaded data and examples used throughout a system. The particular type of output depends upon the domain of the system. Being artists, they create texts and graphics and have to take care of aspects such as a consistent aesthetic appearance or fitting the content to the culture of the system's user. Correspondingly, the tools used also depend on the domain of the system. Either way, most of these tools are specialized to the content format, such as the Adobe Photoshop[2] graphics editor for graphical content or the Qt

[2]https://web.archive.org/web/20171205125120/http://www.adobe.com/de/products/photoshop.html accessed on 5th of December 2017.

Linguist[3] for translation tables. Depending on the system the content designers might also be domain experts contributing domain specific knowledge to the system, such as mathematical formulas or business rules.

Program designers create and maintain the technical side of a software system. This field covers the design of the overall system architecture as well as fine-granular decision such as the names used in the source code. Further, as test engineers they might also write automated tests for checking whether the system behaves as expected, or as tool engineers, they might create tools for making the overall design of the system easier for all roles. When working on a new feature, program designers add, edit, and remove source code. These changes to the source code are often done in so-called integrated development environments (IDEs) which provides a set of integrated development tools in one environment. When working on the overall structure, program designers often use graphical modelling tools that allow them to draw diagrams representing the system structure.

2.4 The Impact of Tools on Cooperation and Collaboration

Agile processes, such as XP, try to tackle the wickedness of software development similar to the way design thinking methodologies tackle wicked problems. Collaboration of multi-disciplinary teams is a key component of Extreme Programming. Thus, XP lists a number of principles and practices to foster this collaboration in social interactions.

However, when it comes to actually working together on the system to be created, the software tools used by team members do not support close collaboration. Every team member uses a specialized tool set to produce artifacts which are particular to their activity. Regarding the concrete artifact to be produced, team members might get feedback in a short amount of time, for example a content designer creating a new icon can see the icon directly in the graphics editor. However, the impression of the icon in the running system might only become available much later when the files representing the icon are merged into the software system. This is similar to the way production in a factory works: individual workers working on optimized stations with their specialized tool. The resulting end product might never be visible to them. Both share the characteristic that there is a long delay between the new artifact produced and a visible change in the resulting system, which might span hours or days.

Such a long delay between one's modifications and an actual change in the software to be created hinders individual team members in assessing the overall state of the software and their impact on it. They work on their local view of the system for an extended period of time. As a result, interdependencies between

[3]https://web.archive.org/web/20171205125218/http://doc.qt.io/qt-5/linguist-translators.html accessed on 5th of December 2017.

modifications, positive as well as negative ones, can only be detected late in the process. For example, translators have to wait for the merge of their translation tables only to find out that a translated text is too long because the interface designer changed the width of some buttons simultaneously. For translators to find such problems, they would have to review their translated texts in the user interface at a later point in time. Further, these long delays can lead to actual deadlocks between two roles with one team not being able to continue working, for example without being able to examine the dimensions of new graphics. Sharing partial results early on would improve this situation.

Agile teams of program designers generally strive for a short roundtrip time between someone's change to the system and a visible change in the system behavior. For example, Extreme Programming proposes having only one branch of source code and just working on separate branches for a few hours maximum. Thereby, all changes to the system always become visible to other team members at least at the end of the day. Although, XP promotes a whole team, this practice only refers explicitly to source code. Other artifacts relevant to the system are not mentioned. In the end, only program designers can effectively modify the system.

Ideally, every team member would contribute to the system directly. This would still allow for experts to work on their tasks with a special tool set, for example a wireframe editor. However, the resulting artifact should directly have an impact on the system, for example the wireframe could directly determine the layout of a user interface without a program designer translating from the wireframe to source code. Thus, the team would work in a *workshop*-like environment in which the final product is at the center and while every team member would work on it using their specialized tools, they would still all contribute to one result. Further, as they would all work in the same room, they can always see the changes made by others and the overall state of the product.

An example of how design teams can implement such a workshop-like environment can be found at Boeing. The team constructing the Boing 777 airplane used a shared 3D model which was always kept up to date with the newest modifications from each team. Further, every team could access it, see the overall state of the airplane design, and examine any interactions between their modifications and the modifications of others (McChrystal 2015).

For software development, there is no need for an additional model as the system to be designed is already a digital artifact and could theoretically directly be accessed by every team member.

3 Learning from Program Designers

Exploratory *programming* environments are based on "the conscious intertwining of system design and implementation" (Sheil 1983). They rely on a variety of properties to support divergent and convergent approaches throughout the design process (Trenouth 1991). However, so far they are based on a very narrow definition of software design as programming. The properties of these environments might actually be generalized to form the conceptual foundation to describe exploratory *software design* environments in which all roles can benefit from these properties. We will first describe the original idea of workflows in exploratory programming environments, the corresponding properties, and illustrate them with two exploratory programming systems: Squeak/Smalltalk (Ingalls et al. 1997) and Lively Kernel (Ingalls et al. 2008; Lincke et al. 2012). We then describe a generalization of this workflow for "exploratory software design environments" and the adapted properties for such systems.

3.1 Exploratory Programming Environments for Program Designers

The idea of exploratory software development originates from the observations that static, linear development processes do not cope well with complex and often-changing requirements. While the process model was not very explicit, the idea was helpful in shaping programming environments which support the iterative and divergent style of programming, which are called exploratory programming environments (Sheil 1983; Trenouth 1991; Sandberg 1988). According to a survey by Trenouth, four properties define such systems (Trenouth 1991):

- *Continuously executable*: The product of the exploration process might not only be the software system but also a greater insight that will inform the future process. Thus, a mere static representation of the software as source code is not desirable. The system to be created should ideally be continuously running and usable.
- *Easily extensible*: Programmers should be able to modify the software easily "without adversely affecting existing behavior".
- *Conveniently explorable*: In order to allow the exploration of design alternative, the environment should support the management of alternatives. It should, consequently, allow programmers to quickly switch between the alternatives.
- *Usefully explainable*: The exploratory programming process aims to allow programmers to understand the problem and design space. As such, the environment should provide means to enable programmers to understand the system, for example through state inspection or visualizations of the dynamic system behavior.

3.2 Case "Desktop Development": Squeak/Smalltalk

Squeak/Smalltalk is an exploratory programming environment (Ingalls et al. 1997; Sandberg 1988). It was designed as a media-authoring and simulation environment.

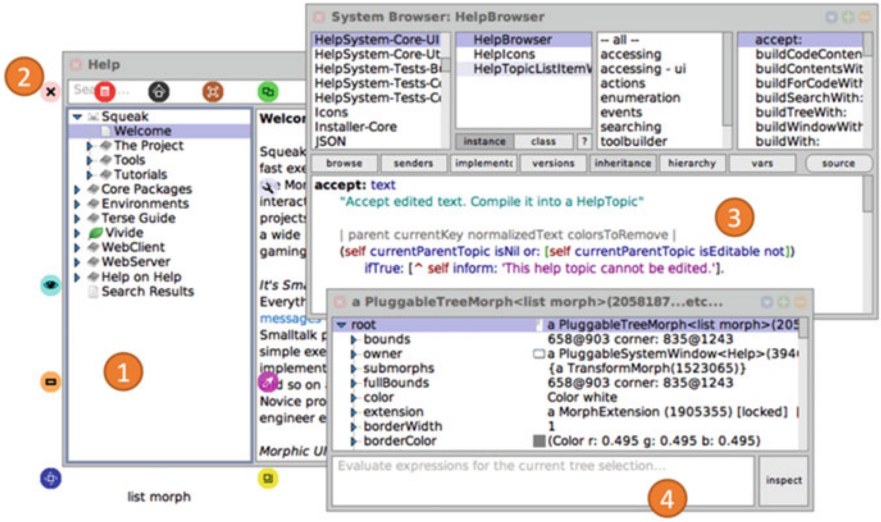


Fig. 3 A screenshot of a list morph on the left (1) with an open halo (2). On the right, the code browser (3) shows parts of the implementation of the help browser tool shown on the left. On the bottom right, an object explorer (4) shows the internal state of the list morph on the left

Several versions and extensions explicitly targeted children exploring ideas and models through the environment.

A fundamental principle of the environment is object-orientation which states that every "thing" in the environment is active in the sense that it has some behavior. This behavior is invoked by sending a message to such an object. For example, sending the message capitalized to the object representing the text "smalltalk" would result in the text calculating a capitalized version of itself which is "Smalltalk". Object-orientation is fundamental for Squeak/Smalltalk as everything is an object. This means all artifacts making up the system, such as source code, pictures, sounds, and layout specifications, are objects.

Squeak/Smalltalk provides special support for the exploratory creation of graphical objects. All visual elements on the screen are so-called *morphs*. A morph can be manipulated through *halos*—a kind of meta-menu allowing access to graphical operations such as resizing or rotation (see Fig. 3). Through the halos, users can also copy a morph and thereby create multiple versions of a morph. Beyond these graphical operations, halos also give users access to some programming facilities per mouse click such as defining the behavior of the morph.

Squeak/Smalltalk is also used as a programming system and thus it provides mature tool support for exploratory programming (see Fig. 3). Squeak/Smalltalk supports the *continuously executable* features as it allows developers to run applications next to their development tools in the same environment. Programmers can further change the system while it is running without any need to restart it. As Squeak/Smalltalk is a class-based object-oriented environment, it is *easily extensible* trough the addition or modification of classes. The support for *convenient* *exploration* is available for source code as well as runtime state. Alternative versions of the source code can be managed on a small scale through local versioning of methods. The state of the system can be versioned by saving the current state of the running system into an image file. When the system is loaded from that file it will be in the exact same state it was before. Finally, the environment provides tools to support the *usefully explainable* feature. With the object explorer and inspector tools, programmers can inspect and manipulate any object. The Squeak/Smalltalk debugger enables programmers to stop the execution of any Smalltalk process and inspect and manipulate the state on the stack.

3.3 Case "Web Development": Lively Kernel

Lively Kernel is another exploratory programming environment (Ingalls et al. 2008; Lincke et al. 2012). As it originates from the Smalltalk tradition of programming systems, it is also object-oriented and exhibits similar tools for exploratory programming. Additionally, it also provides a graphical interface based on morphs and halos.

Lively Kernel, however, allows its users to create the final applications graphical user interface through direct manipulation. After users have assembled their application by combining morphs they can publish their newly assembled graphical object as a *part*. The place where all the published graphical objects are gathered is called the *PartsBin* (see Fig. 4) (Lincke et al. 2012). Other users of Lively Kernel can instantly see a newly published part and create their own copy by dragging a part out of the PartsBin. These copies of the part can then be modified by other users and

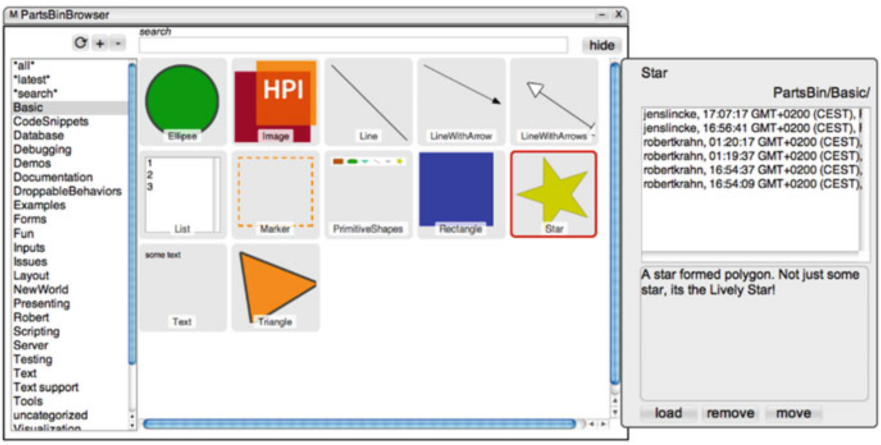


Fig. 4 A screenshot of the parts bin in the Lively Kernel environment. Each graphical element can be dragged out and will create a local copy that can be modified by the local user (Lincke et al. 2012)

republished as a new part. Thereby, all users can quickly make small changes to parts and share them quickly with other team members.

3.4 Towards Environments for Exploratory Software Design

In the original description of exploratory *programming* environments, the features refer to the relation between programmers, source code, and the running system. However, in a more general sense, the features can be applied to any role in the development process to create exploratory *design* environments. Thereby, we move the focus from creating source code to create a running program to creating a variety of artifacts resulting in a software system:

- *Continuously executable*: The software design process should be about creating a working software system that is useful to the user. Thus, everyone participating in the process should be able to execute a current version of the system. This current version should always include a participant's own changes. That means, a content designer creating new sounds should always be able to try these sounds in the environment and interaction designers should always be able to try a new workflow. This property also should hold for the user who should always be able to run a current development version of the system.
- *Easily extensible*: Every team member should be able to easily extend the system in a structured way. A content designer should be able to easily replace or modify content, ideally from within the running system. For example, a graphics designer should be able to modify an icon directly in the running application using graphic editing tools. Consequently, all relevant tools should be included in the environment for every team member. Furthermore, as the design process might unveil new artifacts to be produced, the environment should allow for the easy addition of new tools. Such an environment would even allow members of the design team to change the system in the user's working environmen. In this way, designers can see their effects actual user data and users can immediately see, try, and comment on the modifications.
- *Conveniently explorable*: All design activities within the process profit from an interleaving of divergent and convergent approaches. Convergence is a natural part of the process as there is normally only one current version of the software. In contrast, divergence has to be additionally supported. Thus, the environment should allow versioning and branching for all kinds of artifacts produced in the system. Additionally, switching between versions and comparing versions should also be possible for all artifacts. Ideally, the versioning mechanism is the same for all artifacts including source code.
- *Usefully explainable*: The dynamic nature of a running system affects all artifacts produced. For example, an interface might be layouted differently because the displayed name of a user is too long or the display ration of an icon is distorted because the layout specification changes the border of icons on small screens.

Thus, for all team members to effectively evaluate their modifications, the environment should provide tools for exploring the dynamic version of the artifact in the running system. For example, these tools should allow interface designers to determine which user interactions triggered which transitions in the storyboard so that the system ended up in the current state, or content designers should be able to see scaling parameters for graphics.

4 Identifying Tools and Environments for Whole Team Software Design

The ideal environment containing all tools that might potentially become relevant for any given software project is not possible, due to the variety of domains and constraints for individual projects. However, the idea and the target properties of exploratory software design environments might help in identifying tools and environments which can at least support the collaboration between different roles in a software design team.

4.1 For Individuals: Specific Tools for Specific Tasks

There are a number of tools that integrate the activity of a role with a running instance of the system under design or integrate the artifacts produced by different roles.

Tools have to be created for *content designers* as part of the system when the content is specific to the domain or the system. For example, a system for the automatic assessment of insurance claims might have a dedicated editor for business rules. For more general use cases, generic tools are available which integrate the tool and the running system. For example, the CrowdIn tool[4] allows translators to translate a text interactively directly within the webpage where the text is displayed (see Fig. 5). Further, to make versioning easy for graphics designers and integrate their artifacts with the source code artifacts of the program designers, tools exist which support the versioning of graphics files. An example is the Kactus tool[5] that integrates the graphics editor SketchApp[6] with the versioning tool Git, which is also often used for versioning source code. Thereby, graphic designers and program

[4]https://web.archive.org/web/20171205114216/https://crowdin.com/page/in-context-localization accessed on 5th of December 2017.

[5]https://web.archive.org/web/20171205114339/https://kactus.io/ accessed on 5th of December 2017.

[6]https://web.archive.org/web/20171205124457/https://sketchapp.com/ accessed on 5th of December 2017.

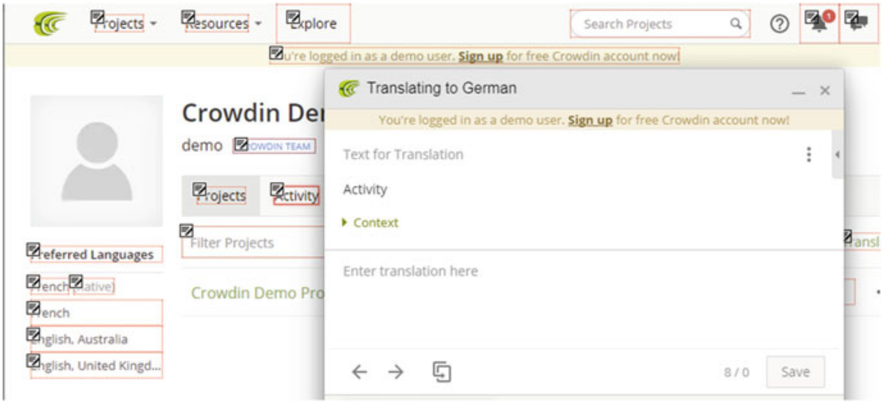


Fig. 5 A screenshot of the CrowdIn tool for translating text of a webpage within the webpage itself

designers can use the same versioning mechanism and see changes from each other throughout the version history.

While *interface design* is often done through mock-ups in graphic editors, it can also be done with tools for creating the actual interface. For example, the Android designer environment includes a layout tool for creating individual screen layouts[7] (see Fig. 6). The iOS development environment further supports the creation of executable storyboards which define the actual transitions between different views in the resulting mobile application.[8] In both cases, the resulting layout files are directly stored in the directory containing the source code of the application and can be shared with the same tools the source code is shared with.

The degree of participation that is possible for *users* again depends mostly on the kind of system to be designed. For specialized systems the user might actually work next door to the design team and might interact with them in person regularly. For a system with a broader target audience this process has to work differently. However, the integration of giving feedback from within the actual context of usage has been improved by several tools. One example is Instabug.[9] In a mobile application containing Instabug, users can add a new suggestion by shaking the phone. The app will stop, create a screenshot, and ask users for further information on what they would have expected in this situation. A research prototype pushed this mechanism further by converting such suggestions directly into stubs and comments in the source code at the appropriate locations (Kato and Goto 2017). Thereby, users can have a very concrete impact on the artifacts making up the system.

---

[7]https://web.archive.org/web/20171205124801/https://developer.android.com/studio/features.html accessed on 5th of December 2017.

[8]https://web.archive.org/web/20171205131025/https://developer.apple.com/xcode/interface-builder/ accessed on 5th of December 2017.

[9]https://web.archive.org/web/20171205131100/https://instabug.com/ accessed on 5th of December 2017.

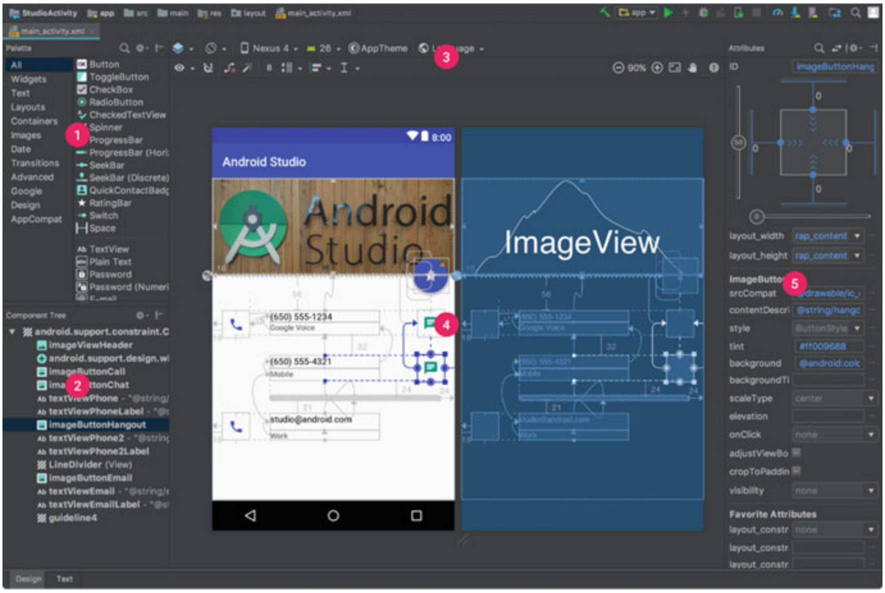


Fig. 6 A screenshot (https://web.archive.org/web/20171205130930/https://developer.android.com/studio/write/layout-editor.html accessed on 5th of December 2017) of the Android layout editor showing new widgets (1), the existing layout tree (2), the toolbar (3), the interactive editor (4), and the property view for one element (5)

Program design is concerned with the behavior of the system, and thus most tools are close to the system in some aspect. Traditional tools separate the modification of source code from the execution of the system, exploratory tools, as described above, integrate the modification of the source code artifacts and the execution of the system (see Figs. 3 and 4).

4.2 *For Teams: Integrated Tool Environments*

For special domains and types of software systems, environments bringing together several roles of the software design team do exist. However, the integration is often based on a thorough understanding of the production processes of the type of software to be created. For example, for a certain type of web application the requirements and efficient development processes are well known. These environments however, would not work well in situations in which requirements are unknown. Alternatively, design tools might be very well integrated for one particular system, which is sometimes done in game development for the design of one particular game.

One example for integrated environments are content management systems (CMS) such as the Drupal system.[10] Content designers, users, interface designers, and program designers are thereby provided the tools to modify or use the system. As exploratory environments, however, such systems lack ways to support versioning or branching for comparing different versions. Further, they are specialized in create-read-update-delete (CRUD) systems which are mostly used for managing and publishing digital artifacts.

Another type of system already integrates many of the relevant tools in one environment: game development environments. Games are complex software systems which require a lot of content design. Consequently, the content and program design are well integrated. An example of such an integration are the development tools for a recent game developed at Nintendo.[11] The integration in their development environment spanned several roles. For example, the interface and program designers were able to see throughout the game world where test users failed most often and could make changes accordingly. For task management, program and interface designers could switch to a task view to see the tasks located next to the relevant location in the world and easily get "get a look at overall completion rates for the game." Further, all content designers were handed the same set of tools: "They created a dedicated software launcher for all the artists to ensure that they were running the same dev[elopment] environment syncing Maya preferences and running automatic tool tests." This focus on integration might be a result of the culture of Nintendo, which the game designers described as: "[. . .] at Nintendo, above all else the most important thing is the fun. This needs to be first and foremost in everyone's mind, regardless of occupation, and they have to tune [in] until the very end to ensure it."[12]

Another environment integrating the activities of several roles is the Home environment which allows the use and modification of productivity tools such as todo lists, e-mail management, or document editing in one environment (Rein et al. 2017). It is based on Squeak/Smalltalk and Vivide (Taeumel et al. 2014) and thus inherits its exploratory properties. However, the Home environment additionally adds user interface elements which make the system usable as an ordinary desktop system. Users can write emails, create todo items, and store them in a hierarchical ordering system similar to a file system. At the same time all tools can directly be modified using the built-in programming tools without any additional setup or any mode changes. This enables users and program designers to work in the same environment. Program designers can make live changes in a user's environment or users demonstrate their desired workflows directly within the environment of a program designer.

[10] https://web.archive.org/web/20171205124956/https://www.drupal.org/ accessed on 5th of December 2017.

[11] https://web.archive.org/web/20171205124841/https://medium.com/@gypsyOtoko/the-final-botw-cedec-session-as-far-as-i-know-is-from-the-engineers-botw-project-management-c30f4e42598e accessed on 5th of December 2017.

[12] https://web.archive.org/web/20171205124841/https://medium.com/@gypsyOtoko/the-final-botw-cedec-session-as-far-as-i-know-is-from-the-engineers-botw-project-management-c30f4e42598e accessed on 5th of December 2017.

5 Conclusion

We have described *exploratory software design environments* as a new perspective on the tools used throughout software development teams consisting of program designers, content designers, user experience designers, and users. Taking inspiration from exploratory programming environments, these environments provide individual team members more direct feedback from the system to be designed, regardless of their own role. Consequently, team members can get an overview of the current state of the system and see the interaction between their modifications and modifications of others. While the creation of one true exploratory software design environment is a wicked problem in itself, individual tools and environments supporting some form of collaboration do exist. By using such tools and environments, teams can grow closer together and create an experience of collaborating while creating a system that brings value to its users.

References

Arnold, J. E. (1956). *Problem solving – A creative approach (National Defense University, Publication No. L57-20)*. Washington, DC: Industrial College of the Armed Forces.

Arnold, J. E. (1959). Creative engineering. In W. J. Clancey (Ed.), *Creative engineering: Promoting innovation by thinking differently* (pp. 59–150). Stanford Digital Repository. http://purl.stanford.edu/jb100vs5754 (Original manuscript 1959).

Beck, K. (2000). *Extreme programming explained: Embrace change*. Boston: Addison-Wesley Professional.

Buchanan, R. (1992). Wicked problems in design thinking. *Design Issues, 8*(2), 5–21.

Conklin, J. (2006). *Dialogue mapping: Building shared understanding of wicked problems*. New York: Wiley.

DeGrace, P., & Stahl, L. (1990). *Wicked problems, righteous solutions*. Upper Saddle River, NJ: Yourdon Press.

Ingalls, D., Kaehler, T., Maloney, J., Wallace, S., & Kay, A. (1997). *Back to the future: The story of squeak, a practical smalltalk written in itself*. ACM SIGPLAN Notices, ACM.

Ingalls, D., Palacz, K., Uhler, S., Taivalsaari, A., & Mikkonen, T. (2008). The lively kernel: A self-supporting system on a web page. In *Proceedings of the Workshop on Self-Sustaining Systems (S3) 2008*. Springer.

Kato, J., & Goto, M. (2017). User-Generated variables: Streamlined interaction design for feature requests and implementations. In *Proceedings of the Programming Experience Workshop (PX/17) 2017*. ACM.

Lincke, J., Krahn, R., Ingalls, D., Röder, M., & Hirschfeld, R. (2012). The lively partsbin – A cloud-based repository for collaborative development of active web content. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS) 2012*.

McChrystal, S. (2015). *Team of teams*. New York: Portfolio/Penguin.

Rein, P., Lincke, J., Ramson, S., Mattis, T., & Hirschfeld, R. (2017). *Living in your programming environment: Towards an environment for exploratory adaptations of productivity tools*. In Proceedings of the Programming Experience Workshop (PX/17.2) 2017. ACM.

Rittel, H., & Webber, M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155–169.

Sandberg, D. W. (1988). Smalltalk and exploratory programming. *ACM SIGPLAN Notices, 23*(10), 85–92.

Sheil, B. (1983). Power tools for programmers. *Datamation Magazine, 29*(2), 131–144.

Taeumel, M., Perscheid, M., Steinert, B., Lincke, J., & Hirschfeld, R. (2014). *Interleaving of modification and use in data-driven tool development*. In Proceedings of the ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming & Software (Onward!) 2014. ACM.

Trenouth, J. (1991). A survey of exploratory software development. *The Computer Journal, 34*(2), 153–163.

“I Know It When I See It”: How Experts and Novices Recognize Good Design

Kesler Tanner and James Landay

Abstract Design novices have limited design experience and typically lack the skills or confidence to create good design, however, they may be able to recognize good design. To assess this ability, 53 novice designers and 52 expert designers participated in an online study where they evaluated a series of websites based on aesthetic appeal using two different modes of comparison. Results show that both experts and novices are able to recognize good design and that novices are able to do so almost as well as experts (76.5% accuracy compared to 81.2%). The greatest determinant of whether a participant would correctly identify a higher-rated design was the difference in the two websites' ground-truth aesthetic ratings. However, expertise and the mode by which the comparison was presented had a significant impact on accuracy (*Keep-the-Best* = 83.6% and *Tournament* = 74.1%).

1 Introduction

While not everyone may be capable of preparing a well-cooked steak, most people may feel they can identify a great steak when eating it, especially if compared with a steak from their local all-you-can-eat buffet restaurant. People may feel similarly about music. Although they may be incapable of composing the next great symphony, their ears can discern between works of great musicians like Mozart and those of the high school rock band practicing down the street.

At some point, however, people's ability to distinguish between good and bad, whether that be food, music or something else, breaks down. The difference in quality becomes too small to be perceived, and people resort to guessing. Experts through training and experience may develop an increased ability to discern and judge quality such that they can still separate items even when they are indistinguishable to a novice. The question remains, however, how big is that difference? At what point does a person's ability to assess difference in quality break down?

K. Tanner (✉) · J. Landay
Stanford University, Stanford, CA, USA
e-mail: keslert@stanford.edu; landay@stanford.edu

C. Meinel, L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation,
https://doi.org/10.1007/978-3-319-97082-0_13

There is also a question of subjectivity. Is there a meaningful scale for judging steak quality, or do people's preferences differ too greatly? How subjective are these orderings?

Like food and music preferences, design is an area in which quality is believed to be highly subjective. Design is also similarly pervasive. Millions of people engage in the design process on a daily basis, creating slide presentations, social media graphics, websites, etc. Some of these people are design experts, trained in best design practices with hundreds of hours of study and experience under their belts, but they are the exception. Most people have limited expertise and understanding of design principles.

Even though most people are not actively participating in the design process, they are exposed to numerous examples of design as they browse the internet, go shopping, and drive on the highway. We hypothesize that, similar to a person's ability to recognize high quality food, the constant exposure people have to design causes them to subconsciously create a basic framework within which to judge design. Even though they are not purposefully training to become experts in design, they can use this framework to recognize good design when they see it. We further hypothesize that while good design is subjective, there is a high level of agreement between people in assessment of design quality.

To test these hypotheses, we built on a study conducted by Reinecke and Gajos (2014) in which they explored design preferences of novices throughout the world by having participants complete an online survey where they rated the visual appeal of websites on a scale of 1–9. We conducted a similar survey in which participants rated a selection of the same websites on a scale of 1–9. Additionally, to rate a second selection of Reinecke and Gajos websites, we used two different comparison methods in which the participants were presented with pairs of websites and asked to select which of the two websites was more visually appealing. While design includes more than visual aesthetic appeal, we chose to focus on this aspect of design since visual appeal has a strong correlation on the perceived usability of something (Hassenzahl 2004), and people that find a design appealing are more tolerant of usability issues (Kurosu and Kashimura 1995). We also collected sufficient demographic information from our participants to separate them into novice and expert categories based on their design expertise.

Based on the analysis of our data, we found that novices are almost as accurate as experts when discerning between the aesthetic qualities of two websites. We also found that while design is indeed subjective, there is also a high degree of agreement about the quality of a visual design. Finally, we found that the mechanism used to present design comparisons has an impact on the overall accuracy and time taken.

The main contributions of this work are: (1) a quantitative assessment showing that visual design quality is not purely subjective due to the high degree of agreement, and (2) empirical data showing that novices are almost as accurate at recognizing good design as experts (within 5%). Because of these primary contributions, this research informs better design tools, and the creation of a meaningful design scale that can be created from more easily obtainable novice comparisons.

2 Related Work

In this section we discuss related work regarding differences between experts and novices, obtaining design feedback from novices, and design intuition.

2.1 Differences Between Experts and Novices

Novices and experts by definition are separated based on their level of expertise. With their higher level of expertise, experts are more familiar with successful design practices and principles. They understand the benefits of parallel prototyping and starting broad with many different ideas before honing in on a final solution, compared to novices who tend to take a "depth first" approach, exploring one design at a time (Cross 2004). As demonstrated by Christiaans and Dorst in their study of junior vs. senior industrial design students, even novices who do recognize the need for seeking outside inspiration and exploring multiple ideas tend to get caught in the information gathering stage and are unable to progress to synthesis (Christiaans and Dorst 1992). On the other hand, experts, with time, develop a repository of solutions. When facing a new problem, they map existing solutions [e.g. design patterns (Duyne et al. 2007; Tidwell 2005)] to new problems in creative ways, whereas novices lack this knowledge of existing solutions and attempt to create a new solution for each new problem (Lloyd and Scott 1994).

While past research has focused primarily on distinguishing the creative abilities of novice as compared to expert designers, we seek to expand this line of research to explore the differences (or lack thereof) between the design quality recognition abilities of these two groups.

2.2 Design Feedback from Novices

Due to the difficulty of obtaining feedback from experts, novices have increasingly been turned to for design feedback. This feedback has taken the form of online task workers using an interface, social media requests, or classroom peers writing a critique. Novice feedback has been found to be helpful to designers, but is perceived to be not as valuable as expert feedback. Research has shown that providing novices with a structure or "scaffolding" with which to provide their feedback helps close the gap between the quality of feedback given by experts and novices (Willett et al. 2012; Xu and Bailey 2012). In fact, Alvin Yuan et al. determined that although an online crowd may seem to lack relevant domain experience, by requiring a non-expert crowd to use a rubric to provide feedback, novice feedback was "rated nearly as valuable as expert feedback" (Yuan et al. 2016). Furthermore, this gap becomes even less significant when timeliness and "clear messag[ing] to a target audience" are the primary concerns of the needed feedback, as opposed to a "range and depth of feedback" (Xu et al. 2015).

Systems such as Voyant (Xu et al. 2014) and CrowdCrit (Luther et al. 2014) demonstrate the validity, effectiveness, and value of design feedback from a non-expert crowd. Such systems are able to eliminate the need to expend social capital to obtain peer critique and the feedback obtained was also determined by experts to approach the quality of peer critique that could be "enthusiastic[ally]" incorporated into design improvements (Luther et al. 2015). While this research clearly shows that novices have the ability to provide meaningful feedback, we sought to delve deeper to validate the underlying assumptions used in this research, including whether novices are able to recognize good design on par with experts and if there is agreement as to what constitutes *good* design.

2.3 Design Intuition

Experts' intuitive design abilities have been the subject of a large body of research. This research breaks down experts' power of intuition into two primary functions: generating alternatives (intuitive speculation) and choosing between these alternatives (intuitive impulse) (Faste 2017). Emphasis is placed on the learned nature of this intuition (Faste 1995; Petitmengin-peugeot 1999), with Cross claiming that this intuition is "honed over time, the ability to make these sorts of qualitative decisions can be considered the designer's systemic ('intuitive') method, through which insight and technical mastery are developed" (Cross 2004). While experts are certainly actively honing their intuitive abilities, we believe that novices may possess a similar intuitive impulse, or the ability to recognize good design without the purposeful honing of this ability.

3 Experiment

We conducted a study to evaluate the effect of design expertise (novice and expert) and comparison mode (*Keep-the-Best* and *Tournament*) on a person's ability to recognize good design. Our goal was to discover to what degree people generally agree upon a website's aesthetic appeal and how novices and experts differed in their perceptions.

3.1 Materials

We used the website snapshots from Reinecke's data set (Reinecke and Gajos 2014), excluding foreign websites and some we felt were overly recognizable (e.g., Boy Scouts of America and Disney World). This resulted in a total of 338 websites. These websites were originally selected by Reinecke to represent a range of colorfulness, visual complexity, and genre.

3.2 Participants

Two hundred and six participants took part in the study, and were found using convenience sampling from sources such as Slack designer communities, NextDoor, Reddit, and Facebook. People were not compensated for their participation, but were told they would be shown how they compared with others upon the study's completion. We then filtered participants to include only those who completed the study within a reasonable time (2 hours), took the survey from within the United States, did not experience any technical difficulties, and professed to have completed the survey to the best of their ability. After this filtering, 118 participants were remaining.

Participants were then separated into three possible groups based on the following two questions: (1) "Do you or have you worked as a design professional?" and (2) "How many years have you worked as a design expert?"

If a participant answered "*No*" to question 1, they were placed in the *Novice* group. If a participant answered "*Yes*" to question 1 and claimed to have 2 or more years of experience, they were placed in the *Expert* group. If a participant answered "*Yes*" to question 1 and claimed to have less than 2 years of experience they were not included in the study. This left us with 53 novice and 52 expert participants. Of these participants, 50 identified as female and 55 identified as male, and the average age was 32.5 years (SD = 11.3, MIN = 19, MAX = 74). For those classified as experts, the average years worked as a design professional was 5.5 (SD = 3.2).

3.3 Apparatus

The study was conducted on s*******.com, a platform we built to conduct design studies. The website and studies were built using React.[1] The server, including hosting and database, was built using Firebase.[2] Images used in the studies were pre-fetched at the start of the study to ensure no delay occurred during the actual

[1]https://facebook.github.io/react/

[2]https://firebase.google.com/

study. Participants took part in the study remotely, using a personal desktop or laptop. The study could not be taken on mobile devices.

3.4 Procedure

Upon arriving at s*******.com, participants began the study by completing a short set of demographic questions (as used for filtering described above). The main part of the study consisted of three tasks: *Rating*, *Keep-the-Best*, and *Tournament*. The order in which a participant completed these tasks was randomized, and the set of websites used in each task were unique to the task.

During the study, a participant viewed 128 distinct websites (*Rating* = 64, *Keep-the-Best* = 32, *Tournament* = 32). To compile the set of images for each task, we used the 1–9 ratings collected from Reinecke's study to provide an average score for each website. We used this score to order the websites from highest rated to lowest rated. We then divided this spectrum into 16 equally sized buckets. From each of the 16 buckets we drew a random image. This process was repeated twice for the *Keep-the-Best* and *Tournament* tasks, and four times for the *Rating* task.

During the *Rating* task (see Fig. 1), users were asked to rate a website based on its aesthetic appeal from 1 (very unappealing) to 9 (very appealing). A website was shown for 500 milliseconds after which it would disappear and the participant would provide a score. In total each participant rated 64 distinct websites.

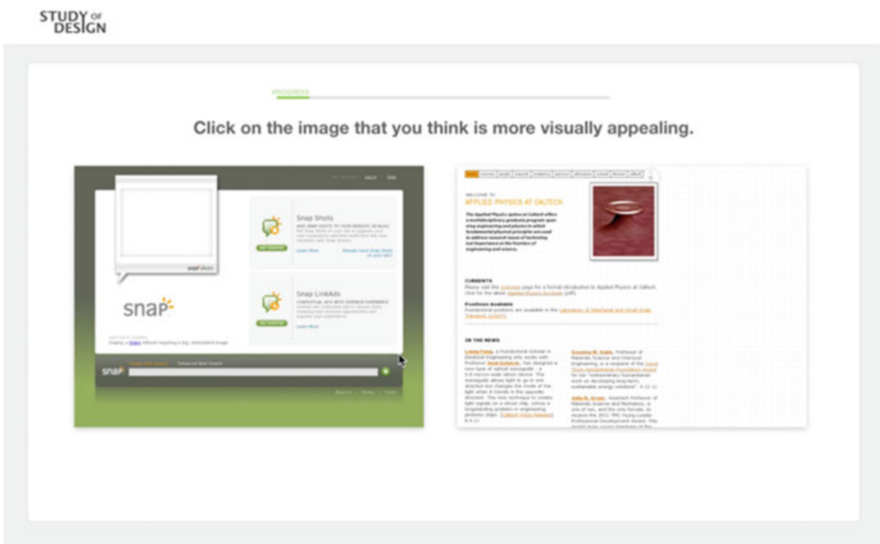


Fig. 1 Interface used during *Rating* task. Participants rated a website on a scale from 1 to 9 to advance to the next decision

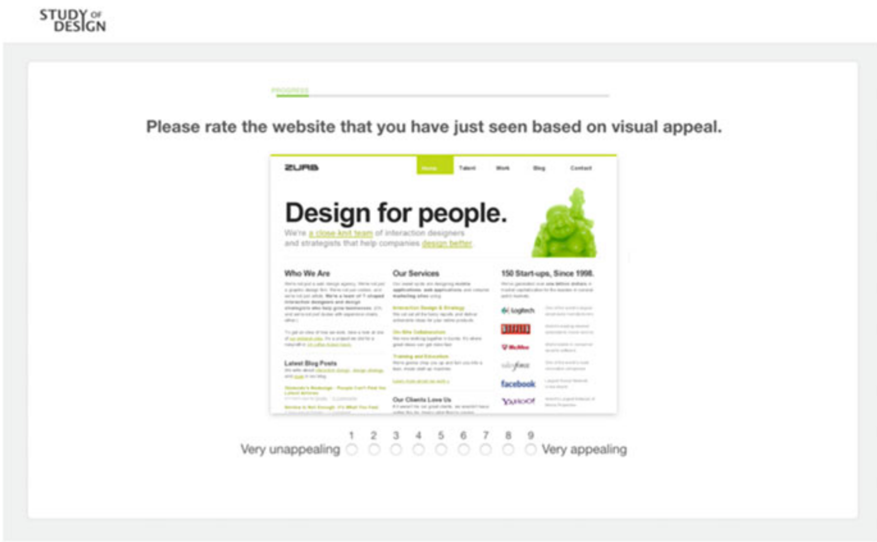


Fig. 2 Interface used during the *Keep-the-Best* and *Tournament* tasks. Participants clicked on the more appealing website to advance to the next decision

The *Keep-the-Best* and *Tournament* comparison tasks were similar. Participants were presented with two websites and asked to select the website that was more visually appealing (see Fig. 2). In the *Keep-the-Best* task, the selected website would be included in the immediate next round and compared against a new website. In the *Tournament* task, the selected website would be added to a "winners pool" and would resurface in future comparisons when all other websites from the current cycle had been compared. Each comparison task included 32 websites, resulting in 31 comparisons for each.

Before beginning each task, the participant was provided with directions and a brief training set using the same four websites (these did not appear in any of the participant's own comparisons or ratings). At the end of each task the participant filled out a NASA TLX form to assess perceived workload. Between each task, participants were given a break (as long as desired) before continuing on to the next section.

At the end of the online survey, participants were asked if they experienced any technical errors, and if they had completed the survey to the best of their ability. On average, the survey took participants 10–15 minutes.

3.5 Data Preparation and Analyses

Our final dataset consisted of 6720 website ratings (3328 expert ratings) and 6510 website comparisons (3224 expert comparisons).

To assess whether a decision was correct during the *Keep-the-Best* and *Tournament* tasks, we needed to establish a system of ground-truth values for the websites. This was originally done using an aggregate of the ratings collected by Reinecke. As an alternative system, we used the data collected during the Rating task, which consisted of 3392 novice and 3328 expert ratings. We found that using the combined expert and novice ratings from our study produced a higher accuracy for both novices and experts during the *Keep-the-Best* and *Tournament* tasks. Since the websites a participant saw in the *Rating* task were unique to that task, a participant's ratings did not improve their own accuracy. Comparing the ordering of websites from Reinecke's and Gajos' data against those obtained in our study yielded a Pearson correlation score of 0.74. We hypothesize that the difference between the two orderings can partially be explained by a difference in 4 years of being collected and the increased percentage of experts in our study.

For each comparison between two websites made during the *Tournament* and *Keep-the-Best* tasks, the following metrics were calculated:

Time: the time in milliseconds from the moment the image appeared to the moment the participant clicked on an image.
Correct: a decision was marked as correct if the participant clicked on the website image that had a higher ground-truth score.
Absolute Difference: the absolute difference between the two websites' scores being compared.

4 Results

Our analysis of variance showed that mode order did not exhibit a main effect. We present our results as a function of *Expertise*, *Mode* and *Absolute Difference*.

4.1 Accuracy

We ran a generalized linear mixed model fit by maximum likelihood to examine the main effects and interaction effects of Figs. 3 and 4, measuring accuracy as a function of *Expertise*, *Mode*, and *Absolute Difference*. There was a significant main effect of *Absolute Difference* on accuracy ($\chi 2$ (1,N = 6510) = 545.04, $p < 0.0001$), as larger absolute differences between websites caused increased accuracy. There was no significant main effect of *Expertise* or *Mode.*

There was an interaction effect between *Absolute Difference* and *Expertise* ($\chi 2$ (1, N = 6510) = 18.68, $p < 0.0001$) as design expertise caused increased accuracy at certain levels of *Absolute Difference.* An expert participant had an average accuracy of 63.4% when the absolute difference between two websites was 0.5 compared to an average accuracy of 62.8% of a novice for the same type of comparison. This

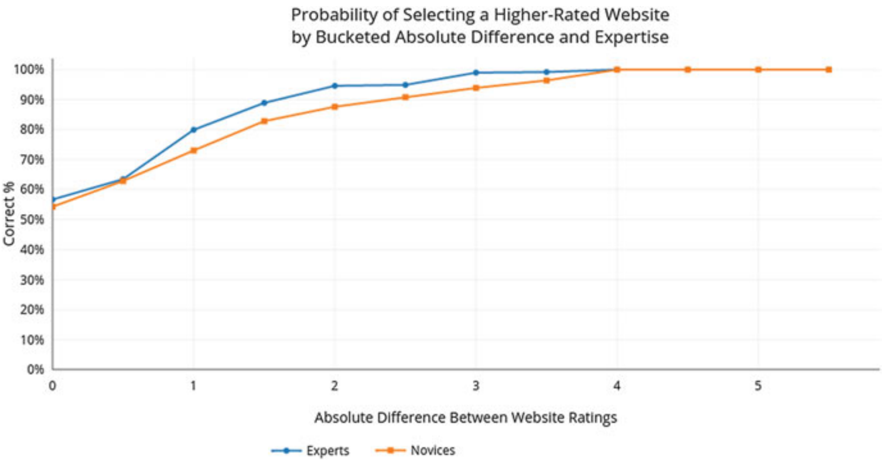


Fig. 3 Probability of selecting a higher-rated website as a function of Absolute Difference and Expertise

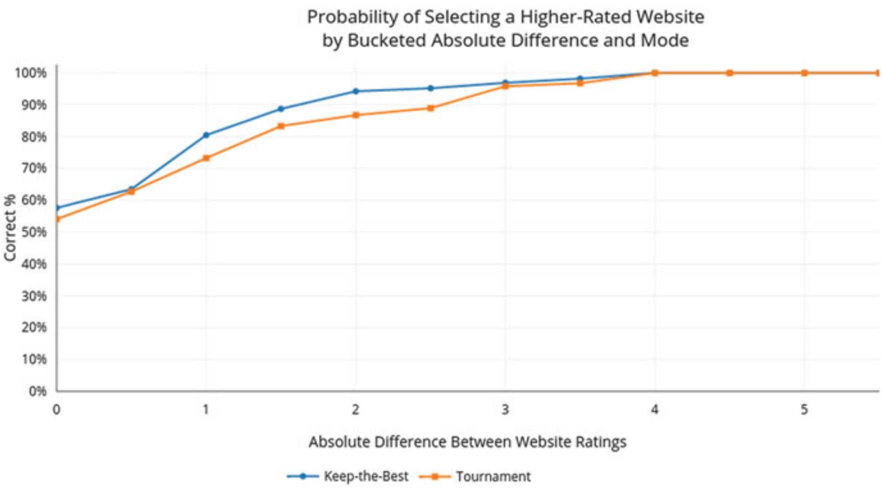


Fig. 4 Probability of selecting a higher-rated website as a function of Absolute Difference and Mode

accuracy increased to 88.9% for an expert and 82.8% for a novice when the absolute difference was ~1.5. When the absolute difference was 4.0 or greater, both experts and novices were 100% accurate. Additional details can be seen in Table 1.

An interaction effect also existed between *Absolute Difference* and *Mode* ($\chi 2$ (1, N = 6510) = 6.91, $p < 0.01$) as the comparison mechanism caused increased accuracy at certain levels of *Absolute Difference*. When comparing two websites during the *Keep-the-Best* task, participants had an average accuracy of 63.5% when the absolute difference between those two websites was ~0.5 compared to an

Table 1 Top shows average accuracy and number of decisions as a function of Expertise and Bucketed Absolute Difference. Bottom shows average accuracy and number of decisions as a function of Mode and Bucketed Absolute Difference

	Bucketed Absolute Difference											
	0	0.5	1*	1.5*	2*	2.5	3*	3.5	4	4.5	5	5.5
Expert												
Accuracy	56.6%	63.4%	79.9%	88.9%	94.6%	94.9%	99.0%	99.2%	100%	100%	100%	NA
# Decisions	362	644	571	530	387	316	198	119	66	29	2	0
Novice												
Accuracy	54.2%	62.8%	73.0%	82.8%	87.6%	90.8%	93.9%	96.4%	100%	100%	100%	100%
# Decisions	371	683	588	512	450	314	180	111	49	17	9	2
	Bucketed Absolute Difference											
	0	0.5	1**	1.5**	2**	2.5	3	3.5	4	4.5	5	5.5
Keep-the-Best												
Accuracy	57.6%	63.5%	80.5%	88.7%	94.2%	95.2%	96.9%	98.2%	100%	100%	100%	100%
# Decisions	269	581	497	496	452	395	258	168	87	41	10	1
Tournament												
Accuracy	54.1%	62.7%	73.3%	83.3%	86.8%	88.9%	95.8%	98.6%	100%	100%	100%	100%
# Decisions	464	662	588	546	385	235	120	62	28	5	1	1

Results where $p < 0.001$ are marked with ** and results where $p < 0.05$ are marked with asterisks average accuracy of 62.7% during the *Tournament* task. This accuracy increased to 88.7% in the *Keep-the-Best* task and 83.3% during the *Tournament* task when the absolute difference was ~1.5. When the absolute difference was 4.0 or greater, participants were 100% accurate during both the *Keep-the-Best* and *Tournament* tasks. Additional details can be seen in Table 1.

Figures 3 and 4 illustrate the connection between absolute difference and accuracy. When the absolute difference between two websites' scores approaches zero, the probability of choosing the higher-rated website approaches 50%. As the absolute difference between two websites' scores increases, the probability of choosing the higher-rated website also increases until it reaches a maximum accuracy of 100%. Table 1 shows that this point of perfect accuracy occurs when the difference between two websites is 4.0.

There was no significant interaction effect between Mode and Expertise, or *Absolute Difference, Mode, and Expertise.*

4.2 Decision Time

We ran a generalized linear mixed effect model to examine main effects, measuring decision time as a function of *Expertise* and *Mode*. In this model, we included the *Rating* task data. There was a significant main effect of *Mode* on decision time ($\chi 2$ (2,N = 13,231) = 98.17, $p < 0.0001$), as decisions made during the *Tournament* task took more time than decisions during the *Keep-the-Best* task. The median decision time during the *Tournament* task was 3234 milliseconds. During the *Keep-the-Best* task, the median decision time was 1965 milliseconds. During the *Rating* task, the median decision time was 2417 milliseconds.

There was no significant main effect of *Expertise* on decision time.

4.3 Nasa TLX

Five of the six NASA TLX categories were used to determine a perceived cognitive load for each task: effort, frustration, mental demand, temporal demand, and performance. Each was rated on a 21-point scale where 1 = Very Low and 21 = Very High. For the performance metric, the labels were adapted to 1 = Failure and 21 = Perfect. Perceived cognitive task load was calculated by averaging the individual scores from each category.

We ran a linear mixed effect model to examine a main effect of perceived cognitive load as seen in Fig. 5. There was a significant main effect of *Mode* on cognitive load (F(2206) = 59.86, $p < 0.001$), as the comparison tasks caused a lower perceived cognitive load than a rating task. Both *Tournament* (T(206) = 10.57, $p < 0.001$) and *Keep-the-Best* (T(206) = 16.01, $p < 0.001$) had a perceived lower cognitive load than *Rating*. *Keep-the-Best* also had a perceived lower cognitive load

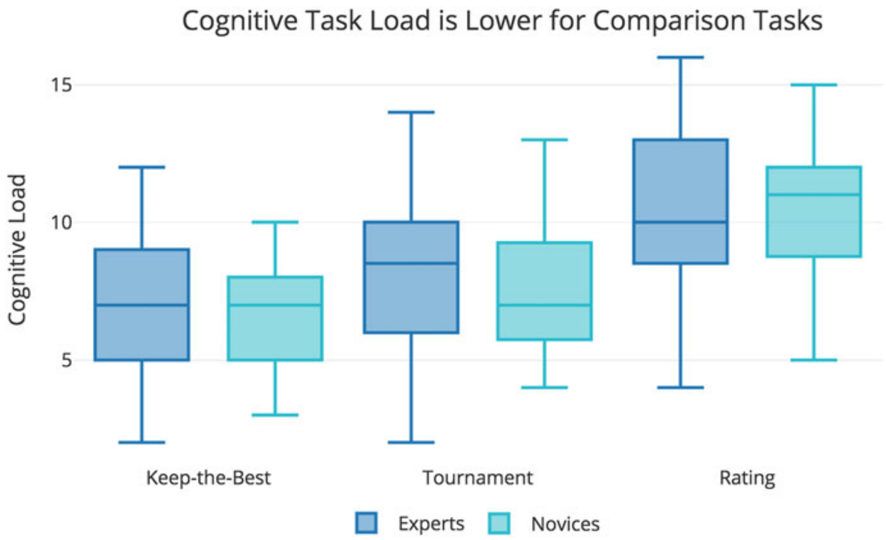


Fig. 5 Perceived cognitive load as self-reported by participants for each of the three tasks

than *Tournament* (T(206) = 5.44, $p < 0.001$). There was no significant effect of *Expertise* on cognitive load.

5 Discussion and Future Work

The goal of this study was to determine if novices are able to accurately recognize good design and how their ability compares to that of an expert. We were able to determine that not only are novices able to recognize good design, but they are able to do so almost as well as experts (within ~5%). While past work (Faste 2017) emphasizes the importance of a learned intuition through the pursuit of professional design experience, our research shows that a participant's natural design intuition is an incredibly powerful ability that can be used without additional training.

5.1 Design Subjectivity and Agreement

The general feeling surrounding design is that its aesthetic quality is purely subjective and cannot be quantified. Our study, however, demonstrated empirically that while design is indeed subjective, there is also a high degree of agreement between both experts and novices, and that some degree of quantification is possible. While our absolute numerical rankings may not be perfect, both they and the comparisons support this correlation. This is further strengthened by the largest signal for whether a study participant would choose a higher-rated website being the absolute difference between those two websites' ratings rather than a participant's level of design expertise or the mode by which the comparison was made. This strengthens the argument that while the aggregated website scores may not be a perfect representation of a website's aesthetic appeal, they are accurate enough to provide a strong indication. While we do agree that design is subjective, it appears that it is highly agreeable and that while a certain design may speak to an individual, on average there is a consensus as to what constitutes good visual design.

5.2 Mode Matters

We found that the *Keep-the-Best* mechanic was superior to the *Tournament* mechanic when asking participants to make comparisons. Not only was it more accurate, but also 50% faster and it required a lower cognitive load. This discovery was particularly surprising based on the anecdotal feedback we received during pilot studies that participants felt significantly more mental baggage during the *Keep-the-Best* task. Participants expressed feeling either that they needed to get rid of designs purely because they'd held on to them for too long, or that they became attached to the design they picked repeatedly and felt they had to continue to choose it to validate past decisions. Numerous participants also cited anxiety regarding how once they eliminated a website during the *Keep-the-Best* task it was gone forever. (Even though websites eliminated during the *Tournament* were also gone forever, no participants cited anxiety relative to this.) The NASA TLX performed as part of the actual study, however, showed that study participants found the *Keep-the-Best* mechanic to be easier than the *Tournament* mechanic, so it is possible that the participants of the pilot study felt more self-conscious with an observer actively watching (and judging) their decisions.

The fact that the *Keep-the-Best* mechanic is not only more accurate and less mentally taxing, but also more closely mimics real-life opportunities (such as parallel prototyping) and applications (including new tools we envision) for comparing designs is promising. That being said, in this study, only two comparison methods were tested. The difference in accuracy, speed, and cognitive load existing between just these two modes suggests that exploring different methods of comparisons could further improve the accuracy, speed, and ease with which websites (or other design tasks) could be examined and compared. Furthermore, based on the interaction effect seen in this study between the absolute difference in ratings between two websites and the mode used, comparison methods could potentially be further refined by incorporating machine learning to optimize the comparison mode based on the absolute difference of the website scores being compared.

5.3 Comparison vs. Rating

We discovered that there were several advantages to comparing two designs and choosing the more visually appealing over rating a design on a 1 to 9 scale. First, participants experienced a lower perceived cognitive load doing a comparison than a rating task. We believe this might be in part because they did not have to maintain a framework for all designs in their minds. To assign a numerical score to a website, a participant needed to have a framework for what constitutes a 3 and how that compares to a 5. Instead, in a comparison, they only needed to compare two websites and determine which was better. The perceived cognitive task load was further decreased in the *Keep-the-Best* mode. This makes sense because in each new round only one new website was introduced, whereas in the *Tournament* mode, a participant saw two new websites each round, essentially doubling the workload.

Second, the median time per decision during the *Keep-the-Best* task was lower than the time per decision during the *Rating* task. This is not substantial when considered as a mere 450 millisecond difference, but collectively over the course of many decisions, it constitutes a 19% speedup. Combined with the lower cognitive load, this type of comparison improves a participant's efficiency while demanding less mental energy.

A third benefit of comparison over rating is it is unnecessary to perform a normalization of user ratings. While one participant might rate websites in the range of 1–6, another might rate the same websites from 4–7. With enough participants, these differences smooth out, however, with comparisons there is no need for a shared framework, as only the order between the two websites matters.

5.4 Future Design Tools

The ability of novice designers to recognize good design to a degree that is on par with design experts is underutilized in current design tools. This idea has surfaced in design tools such as Designscape (O'Donovan et al. 2015), Design Galleries (Marks et al. 1997), and Sketchplore (Todi et al. 2016), but these tools are only tapping the surface of this ability. Often novices start their design task from a blank canvas or a pre-built template. They then iterate on their designs by imagining a change in their mind and then carrying out that change on their canvas.

Instead, we envision tools that focus on a user's ability to recognize good design over their ability to first imagine, and then create it. Such tools could work by rapidly exploring the possibilities in a space and relying on the user as an oracle to validate positive exploration paths. This idea shifts the responsibility of the user from identifying *what might look good*, to identifying *what does look good*. Instead of a user changing the background from white to red, they might instead specify that they want to explore background options. They could then be presented with colors, gradients, images, and patterns that might look good. Although the user only specified background options, such a system should be smart enough to adjust font colors, or text shadows to improve legibility. The user is not required to notice what adjustments are being made, but simply that a particular design is an improvement.

In building these new tools, it is important that differences in proposed alternatives be large enough for comparisons to be valuable. As demonstrated by this study, if designs are too similar in their underlying rating, users will be unable to recognize which is better. In situations where only small, incremental changes are possible, we propose that looping back to an earlier design could confirm that the user is incrementally moving in the right direction, rather than incrementally slipping towards a worse design.

We hypothesize that such tools could also help novices and experts to be less prone to fixation effects (Buxton 2010; Dow and Klemmer 2010), while encouraging good design principles such as parallel prototyping (Dow et al. 2012), iteration (Bogumil 1985; Dow and Klemmer 2010; Hartmann et al. 2006; Salter and Whyte 2004) and seeking external feedback (Tohidi et al. 2006).

5.5 Meaningful Design Scale

Our study demonstrated empirically that while design is indeed subjective, there is also a high degree of agreement between both experts and novices. This finding, combined with the ability of novices as a collective to recognize good design, confirms the validity of the creation of a Meaningful Design Scale (MDS), where any design could be ranked based on aesthetic appeal relative to other previously ranked designs and assigned a numeric score. Creating the MDS would not only provide a framework within which to discuss what is good design in a more quantitative way, but could also provide timely and meaningful feedback to augment the current parallel prototyping feedback loop. This could be accomplished by crowdsourcing a series of novice comparisons to place a design on the scale. While peer feedback asks people similar in ability to assess each other's in-progress work (Kulkarni et al. 2013), our findings suggest that novices could collectively provide useful feedback to experts.

Used in this way, companies could incorporate the MDS as a QA measure for their internal design system (i.e., a landing page can only be put live once it has an MDS score of 8+). We hypothesize that an MDS could be made even more valuable to industry and designers in general by further segmenting novice ratings by demographics (i.e., a women's clothing line could require their landing pages to receive an MDS score of 8+ as assigned by women age 18–34 in the United States). Further research could explore how demographics impact the degree of agreement of perceived aesthetic quality of a design.

6 Conclusion

In this research, we conducted a study to understand how well novices recognize good design. We discovered that novices can recognize good visual design almost on par with experts, and that while design is subjective, there is also a high level of agreement. In addition, we learned that the mode by which a comparison is made has a significant impact on accuracy. We discuss the importance of novice's ability to recognize good design and propose how design tools might better leverage this natural skill, as well as how novices' input could be used to create a meaningful design scale, providing a quantifiable means of discussing design.

Acknowledgments We'd like to thank Jacob O. Wobbrock, Rob Semmens, Kurt Gee and Roy Roy for their assistance with statistics. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect those of any supporter listed above.

References

Bogumil, R. J. (1985). The reflective practitioner: How professionals think in action. *Proceedings of the IEEE* [Internet]. [cited 2017 Sep 15], *73*(4), 845–846. Available from: http://ieeexplore.ieee.org/document/1457478/

Buxton, B. (2010). *Sketching user experiences: Getting the design right and the right design* [Internet]. [cited 2017 Sep 15]. Available from: https://books.google.com/books?hl=en&lr=&id=2vfPxocmLh0C&oi=fnd&pg=PP1&dq=Sketching+User+Experiences:+Getting+the+Design+Right+and+the+Right+Design&ots=06Kuhjl6VO&sig=5VvNqRGGzjzH0ufiVvL8A8UlZXo

Christiaans, H. H. C. M., & Dorst, K. H. (1992). *Cognitive models in industrial design engineering: A protocol study*. American Society of Mechanical Engineers Design Engineering Division.

Cross, N. (2004). Expertise in design: An overview. *Design Studies*. [Internet]. Elsevier; Sep 1 [cited 2017 Sep 15];*25*(5), 427–441. Available from: http://www.sciencedirect.com/science/article/pii/S0142694X04000316

Dow, S., & Klemmer, S. R. (2010). The efficacy of prototyping under time constraints. *Design Thinking*. [Internet]. [cited 2017 Sep 15], pp. 111–128. Available from: http://dl.acm.org/citation.cfm?id=1640260

Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., & Klemmer, S. R. (2012). Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. *Design Thinking Research Studies*. Co-Creation Practice [Internet]. [cited 2017 Sep 15], pp. 127–153. Available from: http://link.springer.com/chapter/10.1007/978-3-642-21643-5_8

Faste, R. (1995). The role of aesthetics in engineering. *Japan Society of Mechanical Engineering Journal*. [Internet]. [cited 2017 Sep 15]. Available from: http://www.fastefoundation.org/publications/the_role_of_aesthetics.pdf

Faste, H. (2017). Intuition in design: Reflections on the iterative aesthetics of form. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* – CHI EA '17 [Internet]. [cited 2017 Sep 15], pp. 3403–3413. Available from: http://dl.acm.org/citation.cfm?id=3025534

Hartmann, B., Klemmer, S., Bernstein, M., Abdulla, L., Burr, B., Robinson-Mosher, A., et al. (2006) Reflective physical prototyping through integrated design, test, and analysis. In *Proceedings of 19th Annual ACM Symposium on User interface Software Technology* [Internet]. [cited 2017 Sep 15], pp. 299–308. Available from: http://dl.acm.org/citation.cfm?id=1166300

Hassenzahl, M. (2004, December 1). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interactions*. [Internet]. L. Erlbaum Associates [cited 2017 Sep 15], *19*(4), 319–349. Available from: http://www.tandfonline.com/doi/abs/10.1207/s15327051hci1904_2

Kulkarni, C., Wei, K. P., Le, H., Chia, D., Papadopoulos, K., & Cheng, J., et al. (2013). Peer and self assessment in massive online classes. *ACM Transactions on Computer-Human Interaction*. [Internet]. [cited 2017 Sep 15], *20*(6), 1–31. Available from: http://link.springer.com/chapter/10.1007/978-3-319-06823-7_9

Kurosu, M., & Kashimura, K. (1995). Creativity of. Apparent usability vs. inherent usability: Experimental analysis on the determinants of the apparent usability. In *Proceedings of ACM Conference on Human Factors in Computer Systems* [Internet]. New York, New York, USA: ACM Press; [cited 2017 Sep 15], 292–293. Available from: http://portal.acm.org/citation.cfm?doid=223355.223680

Lloyd, P., & Scott, P. (1994). Discovering the design problem. *Design Studies*. [Internet]. [cited 2017 Sep 15], *15*(2), 125–140. Available from: http://www.sciencedirect.com/science/article/pii/0142694X94900205

Luther, K., Pavel, A., Wu, W., Tolentino, J., Agrawala, M., & Hartmann, B., et al. (2014) CrowdCrit: Crowdsourcing and aggregating visual design critique. In *Proceedings of Companion Publ. 17th ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW Companion '14* [Internet]. [cited 2017 Sep 15], pp. 21–24. Available from: http://dl.acm.org/citation.cfm?id=2556788

Luther, K., Tolentino, J., Wu, W., Pavel, A., Bailey, B. P., Agrawala, M., et al. (2015) Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of 18th ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW '15* [Internet]. [cited 2017 Sep 15], pp. 473–485. Available from: http://dl.acm.org/citation.cfm?id=2675283

Marks, J., Ruml, W., Ryall, K., Seims, J., Shieber, S., & Andalman, B., et al. (1997). Design galleries: a general approach to setting parameters for computer graphics and animation. In *Proceedings of 24th Annual Conference on Computer Graphics Interactive Technology – SIGGRAPH '97* [Internet]. [cited 2017 Sep 15], pp. 389–400. Available from: http://dl.acm.org/citation.cfm?id=258887

O'Donovan, P., Agarwala, A., & Hertzmann, A. (2015). DesignScape: Design with interactive layout suggestions. In *Proceedings of 33rd Annual Conference on Human Factors in Computer Systems – CHI '15* [Internet]. [cited 2017 Sep 15], pp. 1221–1224. Available from: http://dl.acm.org/citation.cfm?id=2702149

Petitmengin-peugeot, C. (1999). The intuitive experience. *Journal of Consciousness Studies*. [Internet]. [cited 2017 Sep 15], *2*, 43–77. Available from: http://www.ingentaconnect.com/content/imp/jcs/1999/00000006/f0020002/928

Reinecke, K., & Gajos, K. Z. (2014). Quantifying visual preferences around the world. In *Proceedings of 32rd Annual ACM Conference on Human Factors in Computer Systems – CHI '14* [Internet]. ACM Press, New York, USA; [cited 2017 Sep 15], pp. 11–20. Available from: http://dl.acm.org/citation.cfm?doid=2556288.2557052

Salter, A., & Whyte, J. (2004). Serious play: How the world's best companies simulate to innovate [Internet]. *Technovation*. [cited 2017 Sep 15], pp. 277–278. Available from: https://books.google.com/books?hl=en&lr=&id=3f6UdmTaAH0C&oi=fnd&pg=PR9&dq=Serious+Play:+How+the+World's+Best+Companies+Simulate+to+Innovate&ots=RFdJu21jIC&sig=rHiLvl2c91_MH8Zao_nEZmabVhU

Tidwell, J. (2005). *Designing interfaces: Patterns for effective interaction design* [Internet]. OReilly Media Inc. [cited 2017 Sep 19]. p. 352. Available from: https://books.google.com/books?hl=en&lr=&id=5gvOU9X0fu0C&oi=fnd&pg=PR11&dq=Designing+Interfaces:+Patterns+for+Effective+Interaction+Design&ots=sSZ0N7X9VT&sig=eeDGYaJ91-vKulFUKyqqGI2fhBk

Todi, K., Weir, D., & Oulasvirta, A. (2016). Sketchplore: Sketch and explore with a layout optimiser. In *Proceedings of 2016 ACM Conference* [Internet]. [cited 2017 Sep 15]; Available from: http://dl.acm.org/citation.cfm?id=2901817

Tohidi, M., Buxton, W., Baecker, R., & Sellen, A. (2006). Getting the right design and the design right: Testing many is better than one. In *Proceedings of CHI 2006 Conference on Human Factors on Computer Systems* [Internet]. [cited 2017 Sep 15], pp. 1243–1252. Available from: http://dl.acm.org/citation.cfm?id=1124960

Van Duyne, D. K., Landay, J. A., & Hong, J. I. (2007). *The design of sites: Patterns for creating winning web sites* [Internet]. [cited 2017 Sep 19]. Available from: https://books.google.com/books?hl=en&lr=&id=eE2TxLtDsL8C&oi=fnd&pg=PR13&dq=The+Design+of+Sites:+Patterns+for+Creating+Winning+Web+Sites&ots=v9JzHZPmCj&sig=FwQSMbSU8bI_Nno-Gu0-1Wgh9uc

Willett, W., Heer, J., & Agrawala, M. (2012). Strategies for crowdsourcing social data analysis. In *Proceedings of 2012 ACM Annual Conference on Human Factors on Computer Systems – CHI '12* [Internet]. [cited 2017 Sep 15], p. 227. Available from: http://dl.acm.org/citation.cfm?id=2207709

Xu, A., & Bailey, B. P. (2012) What do you think? A case study of benefit, expectation, and interaction in a large online critique community. In *Proceedings of 15th ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW'12* [Internet]. [cited 2017 Sep 15], 295. Available from: http://dl.acm.org/citation.cfm?id=2145252

Xu, A., Huang, S.-W., & Bailey, B. (2014). Voyant: generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of 17th ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW '14* [Internet]. [cited 2017 Sep 15], pp. 1433–1444. Available from: http://dl.acm.org/citation.cfm?id=2531604

Xu, A., Rao, H., Dow, S. P., & Bailey, B. P. (2015). A classroom study of using crowd feedback in the iterative design process. In *Proceedings of 18th ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW '15* [Internet]. [cited 2017 Sep 15], pp. 1637–1648. Available from: http://dl.acm.org/citation.cfm?id=2675140

Yuan, A., Luther, K., Krause, M., Vennix, S. I., Dow, S. P., & Hartmann, B. (2016). Almost an expert: The effects of rubrics and expertise on perceived value of crowdsourced design critiques. In *Proceedings of 19th ACM Conference on Computer Supported Cooperative Work and Social Computing – CSCW '16* [Internet]. [cited 2017 Sep 15], pp. 1003–1015. Available from: http://dl.acm.org/citation.cfm?id=2819953