Understanding Innovation

Hasso Plattner
Christoph Meinel
Larry Leifer *Editors*

Design Thinking Research

Making Design Thinking Foundational

Springer

Understanding Innovation

Series Editors

Christoph Meinel
Larry Leifer

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

Hasso Plattner • Christoph Meinel • Larry Leifer
Editors

Design Thinking Research

Making Design Thinking Foundational

Springer

Editors
Hasso Plattner
Hasso Plattner Institute for Software Systems Engineering
Potsdam
Germany

Christoph Meinel
Hasso Plattner Institute for Software Systems Engineering
Potsdam
Germany

Larry Leifer
Center for Design Research (CDR)
Stanford University
Stanford, CA
USA

ISSN 2197-5752 ISSN 2197-5760 (electronic)
Understanding Innovation
ISBN 978-3-319-19640-4 ISBN 978-3-319-19641-1 (eBook)
DOI 10.1007/978-3-319-19641-1

Library of Congress Control Number: 2015946107

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

It is now more than 10 years ago that the Stanford d.school was officially founded. It did not need much persuasion for me to engage in this unique endeavor, as I was instantly electrified by David Kelly's vision of a hub for innovators on Stanford campus where students and faculty from all departments come together. Looking back, thousands of students with diverse backgrounds at Stanford as well as at the d. school's sister institute at HPI Potsdam have learned and experienced how to tackle wicked problems and complex challenges to come up with innovative, human-centered solutions.

Ever since I came into contact with design thinking, I have been convinced of its innovative power. I have seen design thinking bring about countless examples of unexpected solutions, changes in working cultures, and improvements in team performance. Something, traditional approaches usually simply cannot deliver. They often fall short when it comes to complex and wicked problems which require new approaches and creativity—a skill that is often assumed to be the domain of outstanding individuals. But: Creativity is not just a skill of the chosen few; big ideas do not merely happen; the creative spark does not just ignite on command. That's not how we can find solutions—especially not in a world as complex as ours. Rather, creativity is a process that everyone can implement. And that's what design thinking is all about: it's a framework and method that fosters creative confidence which is foundational for innovation and can ultimately enable everyone to be innovative.

Innovation requires curiosity and an open mind. Design thinking is thinking in terms of opportunities, not restrictions or prohibitions. It is a holistic approach and encourages thinking across boundaries, thereby enabling real and fundamental innovations. Once in contact with design thinking, people experience a sustainable shift in their mindset and how they act and think. There is also a shift in the way people approach challenges. The innovation method design thinking views problems from a human perspective, with the objective of designing innovative products, services, or experiences that are technically feasible, economically viable, and desirable for the target group.

There is a tremendous, growing interest and curiosity in design thinking and a need to scientifically understand its underlying principles. This entails finding out how and why design thinking works and what makes it more successful than other management approaches. Not only do these questions drive research worldwide, but they are also the reason behind my support for the Design Thinking Research Program between the Hasso Plattner Institute in Potsdam, Germany, and Stanford University, USA.

The design thinking method has been successfully taught at both institutions for many years now. The research program and its investigation of the technical, economic, and human factors was the logical consequence of simply teaching the design thinking method. Researchers at both institutions, with diverse backgrounds in disciplines such as engineering, humanities, neurology, or economics, examine how the innovative processes that originate in small, multidisciplinary teams can be improved and further developed in the future. Since the implementation of the Design Thinking Research Program in 2008, dozens of research projects have been conducted; our understanding of this field has advanced and new insights and tools have become available. By taking the understanding of innovation to a new level that is relevant to all disciplines, our research contributes to make design thinking foundational. The findings, however, are not only meant to be discussed in the scientific community. The advances made in design thinking should be made known to the public at large and to all who want and need to drive innovation, be it in companies or society. The publication at hand is our contribution.

Today, design thinking is acknowledged and pursued both in study and in practice even though—or perhaps because—it breaks with traditional approaches. It does so with its focus on human needs, empathy, and team work as well as its valuation of different points of view. Sustainably and deeply impressed by the impact that design thinking has on working culture, creative confidence, and innovative power that I can witness, I am delighted to see the scientific foundation and advancement provided by this research program.

Palo Alto, CA
Winter 2014/2015

Hasso Plattner

Contents

Manifesto: Design Thinking Becomes Foundational

Larry Leifer and Christoph Meinel

Abstract With the integration of design thinking into engineering education, a missing link has been created between the science-focused, context-independent part of engineering and the human society focused, context-dependent aspect. The latter area has long been neglected, partly due to the uncertainty that comes with the unpredictability of human behavior. However, years of design thinking research have improved our understanding of the method's underlying principles. As a result, there has been a breakdown in the skepticism toward design. We can now instrument and quantify design behavior, measure its impact, validate engineering work, and continuously advance our knowledge of design thinking and ourselves. In this paper, we argue that design is ready to become a foundational science for engineering, alongside scientific fields such as physics, chemistry, and biology.

The idea that design thinking is foundational for engineering and, therefore, in an engineering education, is derived from the following five questions. These propositions need to be taken together and in context because in all things human centric (business, design, society, . . .) the knowledge worth having is and must be **context dependent**. This point of view balances the equation with physics and math, whereby the knowledge worth having must be **context independent**.

(1) How Might We Address and Measure the Needs of Society?
While engineering has been described as the application of science and mathematics to the needs of society, up until now we have known and taught our students very little about finding and understanding the needs of society. Thus we have indirectly turned out half-built engineers. These were engineers who could only serve the explicit needs of others without a direct understanding of who had the need and why. With the advent of the design thinking paradigm, this situation has changed.

L. Leifer
Stanford Center for Design Research, Stanford University, Panama Mall 424, Stanford 94305-2232, CA, USA
e-mail: leifer@cdr.stanford.edu

C. Meinel (✉)
Hasso Plattner Institute for Software Systems Engineering, Campus Griebnitzsee, Potsdam, Germany
e-mail: meinel@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_1

Engineers are now taught how to engage with society through empathy training, coping with multiple points of view, actively managing teamwork, and realizing the full potential of product and service prototyping (experimenting). In effect, a critical link has been made between engineering analysis, the science-focused part of the discipline, which is context independent, and engineering design, the human society-focused aspect of the discipline, which is context dependent.

(2) Why Must We Accept That Human Needs Are Context Dependent?
Human behavior is overwhelmingly context dependent (Bandura 1986). This makes the formulation of problems and their solutions difficult.

As engineering faculty and technical managers, most of us did not teach engineering design nor did we attempt to manage design thinking because we either did not understand it and/or because it includes unpredictable human behavior within the system boundary. This lack of understanding and perhaps fear of uncertainty and ambiguity often leads to skepticism and even contempt for the human side of engineering—design.

(3) Are We Ready and Do We Know Enough to Change?
The cumulative work of a global design thinking research community demonstrates our ability to instrument and quantify design behavior. We can measure its impact on corporate team performance. We have started to understand the underlying principles. Though valuable insights have been gained and methods and tools further developed, we are just at the beginning. Biology, too, started as an applied science with scientists such as Darwin gathering knowledge by doing hands-on work. Design is now positioned to follow biology as a foundational science.

However, its application is still, and will always be, the key to a better understanding of design thinking. That's why researchers are striving to gauge and assess its impact. One such an attempt is the recently published study report "Parts Without a Whole? - The Current State of Design Thinking Practice in Organizations" by the project team "Impact by Design Thinking" (Eva Köppen, Holger Rhinow, Jan Schmiedgen, and Christoph Meinel). This study is based on survey results. The researchers not only explored the different forms of design thinking adoption, but also share their valuable insights derived from years of profound research on design thinking in organizations (Schmiedgen et al. 2015).

The benefits of their work can be seen at *thisisdesignthinking.net*. This website launched by the research team showcases interesting stories from companies working with design thinking and publishes interviews with experts and practitioners. By drawing a colorful picture of the manifold design thinking activities going on today, the website serves the research community as well as coaches, practitioners and students.

Many more projects pave the way for making design a foundational science.

(4) Why Is This Breakthrough Happening Now?
While design thinking is practiced universally in varying degrees, it is the unique combination of engineering (especially IT), economics, anthropology, psychology, neuroscience and design research that is making it foundational (Mabogunje et al. 2015). Breakthroughs at Stanford University cannot be separated from the

university's location in Silicon Valley and the community's impact. Breakthroughs at the Hasso Plattner Institute at the University of Potsdam cannot be separated from its inspiring founder and the vibrancy of the Berlin/Brandenburg region.

Both universities, and in particular their schools of design thinking at Stanford and Potsdam, and other pioneering universities around the world attract companies and organizations eager to apply new approaches to their challenges and projects. They seek to launch change processes and find new inspiration for their work. On the other hand, d.school graduates have an open mind, are full of ideas and enthusiasm for innovation with an appetite for new solutions that yield better services, products, and even fundamentally better societies. These graduates transport a new spirit to their employers, multiplying it, and implement their own projects that stem from a deep understanding of people's needs.

Frederick Terman, an early Dean of Engineering at Stanford (and widely recognized as the "father of Silicon Valley") re-conceptualized the role of the university as follows:

> Universities are rapidly developing into more than mere places of learning. They are becoming major economic influences in the nation's industrial life, affecting the location of industry, population growth, and the character of communities. Universities are in brief a natural resource just as are raw materials, transportation, climate, etc.

Universities are now the knowledge creation engines of society, largely replacing industry R&D for radical breakthroughs. They accelerate the creation of new technologies, new ventures, new markets, and new sources and targets for capital formation.

(5) Is It Time to Professionalize Design Thinking?
To build on our understanding of innovation and the role of design thinking behavior, including supporting brain research (Donald 1991), we propose to move forward with the professionalization of design thinking. Imagine that in time there will be professional schools of design thinking, much as we have schools of engineering, schools of medicine, and schools of business. Imagine too the emergence of pan-disciplinary doctoral programs in design thinking practice.

The first d.school at Stanford started in a garage on the outskirts of the Stanford campus. Hasso Plattner, an early sponsor of the design thinking activities at Stanford, recognized the potential as did the university itself. Ever since, thousands of students of all disciplines have been studying and practicing design thinking. It did not take long to transfer the successful concept to the Hasso Plattner Institute in Potsdam—adapting it to the specific context there. Many universities worldwide approach the two institutes in order to implement a similar school following the "d. school" model. Recognizing the value of such a training facility they strive to create the breeding ground for innovation in their own region. In a cooperative effort, schools of design thinking have been established, for example in Paris, Kuala Lumpur, and Beijing, each one with cultural adaptations based on its unique context. More and more governments and universities plan to establish their own schools of design thinking and integrate design thinking into their curriculum. The design-paradigm is worth your attention.

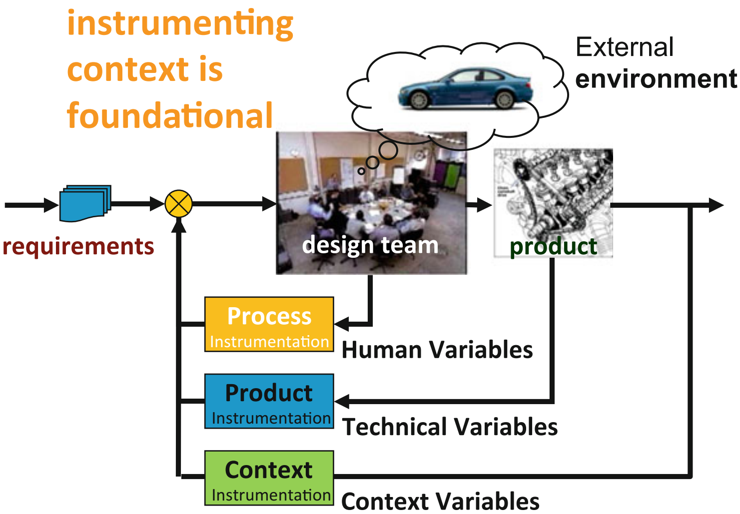


Fig. 1 We build upon advances in design behavior research and corresponding brain research. We quantify action and outcomes. We have a mature measure of predictive power. As a profession, **design thinking is foundational** and can be understood scientifically

Now that we have the roots of the scientific comprehension of design thinking we can expect to continuously improve our understanding of ourselves. We can additionally discover new practices and disseminate these practices through publications, simulations, and emulations. This understanding, which derives from the study of human-human interaction with IT augmentation is embodied in the nature of language—verbal and gestural. It is foundational (Fig. 1).

References

Bandura A (1986) Social foundations of thought and action: a social cognitive theory. Prentice Hall, Englewood Cliffs, NJ

Donald M (1991) Origins of the modern mind: three stages in the evolution of culture and cognition. Harvard University Press, Cambridge, MA

Mabogunje A, Sonalkar N, Leifer L (2015) Design thinking: a new foundational science for engineering. In: To appear in Proceedings of the Mudd design workshop, Harvey Mudd College, Claremont, California, 28–30 May 2015

Schmiedgen J, Rhinow H, Köppen E, Meinel C (2015) Parts without a whole?—the current state of design thinking practice in organizations, Technischer report. Study report, Hasso-Plattner-Institute, Potsdam

Introduction: The HPI-Stanford Design Thinking Research Program

Claudia Koch, Christoph Meinel, and Larry Leifer

Since 2008 scientists from the Hasso Plattner Institute for Software Systems Engineering in Potsdam, Germany, and from Stanford University, USA, have engaged in the joint Design Thinking Research Program, financed and supported by the Hasso Plattner Foundation.

1 Program Vision and Priorities

The multidisciplinary research teams of the HPI-Stanford Design Thinking Research Program scientifically investigate innovation and design thinking in all its holistic dimensions. With backgrounds in engineering, design, humanities or social sciences, team members strive to gain a deep understanding of the underlying principles and, consequently, how and why the innovation method of design thinking works and fails. But the aim is not only to advance design thinking theory and knowledge within the research community. Instead the program seeks to ultimately improve design practice and education by gathering scientific evidence to support design activities.

Applying rigorous academic methods, the researchers study, for example, the complex interaction between members of multi-disciplinary teams. An important feature of this domain is the necessity of creative collaboration across spatial, temporal, and cultural boundaries. The researchers design, develop and evaluate innovative tools and methods that support teams in their creative work. Researchers pursue the common question of why structures of successful design thinking teams differ substantially from traditional corporate structures and how design thinking

C. Koch (✉) • C. Meinel
Hasso Plattner Institute for Software Systems Engineering, Campus Griebnitzsee, 14482 Potsdam, Germany
e-mail: claudia.koch@hpi.de; meinel@hpi.de

L. Leifer
Stanford Center for Design Research, Stanford University, Panama Mall 424, Stanford 94305-2232, CA, USA
e-mail: leifer@cdr.stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_2

methods mesh with traditional engineering and management approaches. Beyond a mere descriptive understanding, the goal of this program is to develop metrics that allow assessment and prediction of team performance in order to facilitate real-time management of how teams work.

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 considered especially important to assess the impact and needed transformations of design thinking in organizations. Projects that set new research priorities in the design thinking research domain are favorably funded. The project selection is further based on intellectual merit and evidence of open collaboration.

Special interest lies in the following guiding 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 really work to create the right innovation at the right time? How do they fail?

2 Road Map Through This Book

Divided into four parts, this book compiles the results of the research projects of the 6th program year, covering multifaceted features and aspects of design thinking.

Team collaboration is the basis of design thinking and thus of significant importance in the program's research agenda. Different "***Tools and Techniques for Improved Team Interaction***" are presented in the first part of this book. These tools and techniques, which have been investigated and developed in the program, range from the Tele-Board for remote collaboration, to IDN for visual diagnostics, to Google glasses in health applications. A technique for physical interaction and a discussion system to leverage geographic diversity in massive online classes are also introduced.

The chapters in Part II all focus on ***"Creativity and Creative Confidence"***— central factors when engaging in design processes. The research projects range from neurological studies on sustainable creativity trainings to the influence of spatial factors on creativity and how creative confidence can be augmented with the help of prototyping tools.

A question that is often asked by people who engage in design thinking, or who intend to do so, is: what is the actual impact of design thinking? Part III ***"Measuring Design Thinking"*** contains a study giving insight into how the impact of design thinking is measured in organizations. Furthermore, metrics for creative behavior are presented in this section.

Part IV, the final section of the book focuses on "***Documentation and Information Transfer in Design Thinking Processes".*** Specifically, this section addresses how to recover previous states in programming processes or how to ensure information transfer between design teams and engineers by recovering innovation paths. It also introduces an approach of special topic coaches sharing their experiences and knowledge. The Tele-Board MED project shows how digital documentation can support design work.

2.1 Part I: Tools and Techniques for Improved Team Interaction

Remote collaboration processes require digital tools supporting work over distances. Transferring physical artifacts to the digital world and facilitating their flexible usage are common cases in remote settings. In "**Globalized Design Thinking: Bridging the Gap Between Analog and Digital for Browser-Based Remote Collaboration**" Matthias Wenzel, Lutz Gericke, Christoph Thiele and Christoph Meinel present new developments that they have integrated into their remote collaboration software system Tele-Board to support such use cases. The authors describe a software tool for automatically digitizing analog whiteboard artifacts that can then be used in remote settings on a shared virtual whiteboard surface via Tele-Board or a web browser-based virtual whiteboard application. This tool allows shared real-time collaboration and, by closing the media gap, lowers the hurdle of switching from analog co-located to digital remote working modes.

The tools currently available for effective team coaching are limited to heuristics derived from either experienced design thinking professionals or clinical psychology practitioners. In "**Visual Diagnostics for Design Thinking Teams**" Neeraj Sonalkar, Ade Mabogunje, Gina Pai, Aparna Krishnan, and Bernard Roth aim to improve this current situation by providing design thinking managers, coaches and instructors a scientifically validated tool for augmenting design team performance. They present the development of a software tool called the IDN Tool, based on the Interaction Dynamics Notation. This tool analyzes team interactions and diagnoses patterns of behavior that influence design outcomes.

In "**Design Thinking Health: Telepresence for Remote Teams with Mobile Augmented Reality**" Lauren Aquino Shluzas, Gabriel Aldaz, and Larry Leifer examine the capabilities and boundaries of a hands-free mobile augmented reality (AR) system for distributed healthcare. They use a developer version of the Google Glass™ head-mounted display to develop software applications to enable remote connectivity in the healthcare field. With this technology, the team characterizes system usage, data integration, and data visualization capabilities. Further, they conduct a series of pilot studies involving medical scenarios. This chapter discusses the need for a AR head-mounted display to improve chronic wound care photography and to facilitate surgical interventions. The authors provide an overview of the system architecture used in this research, and highlight future applications of AR systems for improved clinical care.

Massive online classes are global and diverse. How can we harness this diversity to improve engagement and learning? Currently, despite high enrollment, the interaction between students is minimal. In fact, most students participate without reciprocal action with others. This isolation is particularly disappointing given that a global community is a main draw of online classes. "**Talkabout: Making Distance Matter with Small Groups in Massive Classes**" by Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S. Bernstein, and Scott Klemmer illustrates the potential of leveraging geographic diversity in massive online classes. The authors connect students from around the world through small-group video discussions. In this work, the authors challenge the view that online classes are useful only when in-person classes are unavailable. Instead, they describe how diverse online classrooms can create benefits that are largely unavailable in a traditional classroom.

Over the past 2 years, David Sirkin, Brian Mok, Stephen Yang, Rohan Maheshwari, and Wendy Ju have followed an improvisational approach to physical interaction in design research. This approach emphasizes the use of exploratory lab and field experiments as a way to (a) challenge our assumptions about how people might interact with expressive objects such as robots and active spaces, (b) appraise the performance of our prototypes of these technologies, and (c) build frameworks to understand users' mental models and develop new insights into interaction. They have focused, in particular, on creating environments—whether in public settings or recreated in their workspace—where they can observe people's natural reactions to subtle changes in their routine, or to more explicit provocations designed to understand their deeper thought processes. "**Improving Design Thinking Through Collaborative Improvisation—How We Use Naturalistic Environments to Test Design Ideas**" describes how the team designs and runs experiments to evaluate how people interact with robots built from everyday objects.

2.2 Part II: Creativity and Creative Confidence

Grace Hawthorne, Manish Saggar, Eve-Marie Quintin, Nick Bott, Eliza Keinitz, Ning Liu, Yin-Hsuan Chien, Daniel Hong, Adam Royalty, and Allan L. Reiss contribute "**Designing a Creativity Assessment Tool for the 21st century: Preliminary Results and Insights from Developing a Design-Thinking Based Assessment of Creative Capacity**". In order to assess a person's creative capacity in real-world situations, they propose a novel Design Thinking Creativity Test (DTCT). The DTCT is based on the design thinking principles and can serve as an assessment to reflect problem-solving needs of the twenty-first century. In particular, the DTCT emphasizes assessment of case-based skills to directly measure an individual's application of creativity during an innovation event. The authors provide a thorough background on the tools that already exist to assess creative capacity. We learn how the authors' approach advances the current state of the art. They also share challenges faced in collecting and analyzing DTCT data, along with proposed solutions. Lastly, they provide their hypotheses and preliminary insights regarding DTCT's ability to capture applied creative capacity.

A flexible work environment enables innovation teams to ideate, create and design. "**Innovation in Creative Environments: Understanding and Measuring the Influence of Spatial Effects on Design Thinking-Teams**" by Claudia Nicolai, Marie Klooker, Dora Panayotova, Daniela Hüsam, and Ulrich Weinberg investigates the impact of creative environments on team wellbeing and performance. The creative environment is shown to be drawn from the perceptions, feelings and interactions of individual team members with their respective environment. The study introduces a new qualitative method, cultural probes, as an empirical instrument. The results of this study indicate that innovation teams need access to flexible spatial environments to fulfill their innovation process tasks (performance-oriented perspective), but also need different working zones to foster their team-wellbeing related activities (team-oriented perspective).

Can we enable anyone to create anything? The prototyping tools of a rising "maker movement" are enabling the next generation of artists, designers, educators, and engineers to bootstrap from napkin sketch to functional prototype. However for technical novices, the process of including electronic components in prototypes can hamper the creative process with technical details. Software and electronic modules can reduce the amount of work a designer must perform in order to express an idea, by condensing the number of choices into a physical and cognitive "chunk." What are the core building blocks that might make up electronics toolkits of the future, and what are the key affordances? In "**Building Blocks of the Maker Movement: Modularity Enhances Creative Confidence during Prototyping**" Joel Sadler, Lauren Shluzas, Paulo Blikstein, and Riitta Katila present the idea that modularity, the ability to freely recombine elements, is a key affordance for novice prototyping with electronics. They present the results of a creative prototyping experiment that explores how tool modularity influences the creative design process, highlighting its importance in creative prototyping.

2.3 Part III: Measuring Design Thinking

"**Measuring the Impact of Design Thinking**" by Jan Schmiedgen, Lea Spille, Eva Köppen, Holger Rhinow, and Christoph Meinel focuses on how organizations measure the impact of design thinking. This article is based on a quantitative survey that was supplemented with qualitative interviews with experienced design thinkers. Even though the majority of respondents perceive that design thinking has an impact of some kind, only a small number of them attempt to measure it in some form. Those who do not measure the impact of design thinking often do not know how or lack the resources to do so. The metrics of those who do measure the impact vary considerably, but customer feedback and satisfaction is a recurring theme. The authors propose that traditional means of performance measurements are often ill-suited for evaluating the impact of design thinking. They conclude with a promising industry example of how traditional measures are used to gauge overall performance and a story-based approach to capture the role of design thinking.

The creative behaviors that underpin design thinking are also difficult to measure. This is problematic because people who have a desire to practice design thinking in an organizational context are often assessed solely on their ability to perform via traditional metrics. Therefore they have less incentive to work in a creative way that crosses the boundaries of these traditional metrics. In order for organizations to fully support and incentivize design thinking, they must measure creative behaviors as much as they do execution behaviors. The chapter "**Developing Design Thinking Metrics as a Driver of Creative Innovation**" by Adam Royalty and Bernard Roth highlights a suite of initial metrics that have arisen from research on d.school alumni and organizations who apply design thinking as a core driver of their innovation strategy.

2.4 Part IV: Documentation and Information Transfer in Design Thinking Processes

Design teams have different problems and needs during their projects. For the teams' coaches this can mean a high workload due to familiarizing themselves with new topics and preparing different coaching sessions for the teams. On the other hand, past teams might have experienced similar problems, and the experiences and solutions of members from such teams could be valuable for current teams. In "**Experience and Knowledge Transfer through Special Topic Coaching Sessions,**" Franziska Häger, Thomas Kowark, and Matthias Uflacker present the concept of special topic coaches that allows former members of design teams to provide additional coaching sessions to current teams. This concept makes it possible for experienced members of design teams to share their knowledge and experiences with a current team by preparing and running a coaching session. The researchers implemented the concept with the help of a coaching seminar parallel to one of their design engineering courses. They present their evaluation results and the changes they adapt to the next iteration of the seminar as well as their ideas of how to adopt this concept in a company context.

Documentation is a field of active research for the community of both design thinkers and medical practitioners. One major challenge is to combine the beneficial features of analogue and digital documentation. Since documentation needs are quite similar in the fields of design thinking and behavior psychotherapy, an intense collaboration has emerged between these disciplines. The design thinking tool Tele-Board has been adapted for documentation purposes in behavior psychotherapy and yielded a first medical application of the new tool Tele-Board MED. In the course of tool adaptation, additional features have been developed such as an automatic protocol function. These new features are not only useful for therapists but beneficial for design thinkers as well. In "**Smart Documentation with Tele-Board MED**" Julia von Thienen, Anja Perlich, Johannes Eschrig, and Christoph Meinel explain why collaborative work on documentation tools is particularly promising at the intersection of design thinking and behavior psychotherapy.

Programming stands in the focus of the chapter "**Preserving Access to Previous System States in the Lively Kernel**" by Lauritz Thamsen, Bastian Steinert, and Robert Hirschfeld. In programming systems such as the Lively Kernel, programmers construct applications from objects. Dedicated tools make it possible to manipulate the state and behavior of objects at runtime. However, when programmers make mistakes in such programming systems, they need to undo the effects of their actions. Yet, recovering previous states is often error-prone and time-consuming. This report presents an approach to object versioning for systems such as the Lively Kernel. Access to previous versions of objects is preserved using version-aware references. These references can be resolved to multiple versions of objects and, thereby, allow reestablishing preserved states of the system. The authors present a design based on proxies and an implementation in JavaScript.

Companies implement innovation processes or outsource them to external consulting companies to gain a competitive business advantage. However, as innovators and engineers are seldom the same people, it is inevitable that innovation projects will be documented prior to a subsequent information handover. In practical application this information handover seldom goes smoothly due to missing, incomplete, or non-traceable documentation of innovation projects. The situation becomes even worse when important design rationales, design paths and design alternatives become no longer retrievable. The missing information may lead to an innovation that was not intended by the innovators. The retrieval of design paths, design rationales, and design alternatives requires high manual effort—if it is even possible at all. In "**Connecting Designing and Engineering Activities III**" Thomas Beyhl and Holger Giese present a recovery approach that eases the retrieval of design artifacts by recovering design paths before the actual retrieval happens. In order to do this the authors employ a recovery approach using recovery modules that implement knowledge extraction procedures and recovery algorithms. The authors evaluate their recovery approach using inventory documentation collected in educational design thinking settings.

3 Summary

These articles contribute to a better understanding of innovation and design thinking in particular. Many years of extensive research have yielded valuable insights on why and how this method works. The researchers found metrics, developed models and conducted studies and thereby laid the groundwork to make design thinking foundational.

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. We would be delighted to get in contact with our readers for further discussion and an exchange of ideas. We invite you to visit our websites. At www.hpi.de/dtrp you will find the latest information on past and present research conducted within our program. Learn more about all projects and the researchers behind them. Beyond that, the Electronic Colloquium on Design Thinking Research (ECDTR, http://ecdtr.hpi.de) is an ideal forum for the rapid and widespread exchange of ideas, methods, and results in design thinking research. The purpose of this forum is to use electronic media for scientific communication and discussions in the design thinking research community. The ECDTR welcomes papers, short articles and surveys.

Last but not least, the new website thisisdesignthinking.net offers an easily accessible overview of current developments in design thinking. This pool of examples and interviews, enriched with scientific explanations, helps localize all existing expressions of design thinking, including their advantages and downsides. Here, practitioners searching for advice on design thinking have an opportunity to meet their "corporate twin." This takes the form of analogous organizations, which have the potential to assume a role model function. For educators, the website serves as a source of inspiration for recharging their teaching materials, explanatory models and perspectives on current problems in design thinking practice. Please get in touch with us to share your experiences and stories via thisisdesignthinking@hpi.de.

We invite and encourage you to engage in dialogue with us on your ideas, questions, experiences and insights. May this publication serve you as a deep-dive into design thinking tools, methods and metrics. We hope you enjoy the read and that this book becomes a source of inspiration for your own work.

Part I
Tools and Techniques for Improved Team Interaction

Globalized Design Thinking: Bridging the Gap Between Analog and Digital for Browser-Based Remote Collaboration

Matthias Wenzel, Lutz Gericke, Christoph Thiele, and Christoph Meinel

Abstract Remote collaboration processes require digital tools supporting work over distances. Transferring physical artifacts to the digital world and facilitating their flexible usage are common cases in remote settings.

In this article, we present new developments that we have integrated into our remote collaboration software system Tele-Board supporting such use cases. We describe a software tool for automatically digitizing analog whiteboard artifacts that can then be used in remote settings on a shared virtual whiteboard surface via Tele-Board. Additionally, we show a web browser-based virtual whiteboard application. It allows shared real-time collaboration in a web browser as part of the Tele-Board system. The application makes use of modern web technologies and does not require any browser plugins. This way, it can be used equally on a multitude of hardware, especially mobile devices.

By closing the media gap, our tools lower the hurdle of switching from analog co-located to digital remote working modes. Once arrived in the digital world, our browser-based approach takes account of increased hardware diversity and allows an easy and flexible participation in remote collaboration settings.

1 Introduction

Team-based working modes are an integral part of Design Thinking. Those teams, especially in corporate environments, are increasingly distributed between locations over the globe. The question of how this remote collaboration can be accomplished by Design Thinking teams has been our research focus in recent years (Gumienny et al. 2012a, b; Gericke et al. 2012a). A result of this research was the design and implementation of the remote collaboration tool Tele-Board.

M. Wenzel (✉) • L. Gericke • C. Thiele • C. Meinel
Hasso Plattner Institute for Software Systems Engineering, Campus Griebnitzsee, 14482 Potsdam, Germany
e-mail: matthias.wenzel@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_3

During collaboration, the digital tool is not only a plain functional instrument. In fact, with its usability and acceptance it is a relevant factor for the teams' success. Thus, making the tool an important player among the members of a team requires its interplay with the team's working situation.

There are two aspects that we focused on during the usage-driven iterative development of Tele-Board and its support for distributed Design Thinking working modes.

Mobile devices such as smartphones and tablets play an increasingly important role in today's working environments. In order to participate in remote collaboration processes supported by Tele-Board, our system must be fully accessible by those devices. Current technologies utilized by our system make it difficult to use Tele-Board in mobile scenarios.

A second point affects the decision of using Tele-Board during Design Thinking activity in general. To date, this decision has to be taken right in the beginning of the collaborative process. Starting with analog tools in an early phase of the process and later changing to a digital mode, which is supported by Tele-Board, is difficult. So far, there was no efficient way of transferring real world artifacts, such as sticky notes or handwriting on a whiteboard to the digital world of Tele-Board. Actually, this resulted in an all-or-nothing decision of whether to use our tool or not. Switching from analog to digital was a manual process and therefore a tedious undertaking.

In this chapter, we present solutions for broadening the system's availability on mobile devices as well as increasing its flexibility of usage within the collaboration process.

We describe a web-based re-implementation of the Tele-Board whiteboard component. This application is now completely executable within web browsers without additional software requirements. Browsers are nowadays available on a broad range of devices from traditional desktop computers to mobile devices such as phones. This extends the availability of Tele-Board to a multitude of different platforms. Furthermore, we present our newly developed software tool for automatically digitizing analog whiteboard artifacts that can then be seamlessly used in Tele-Board. This allows using originally analog content in distributed settings as well.

2 Overview of the Tele-Board System

Tele-Board is a digital whiteboard system that allows creative teams to work together over geographical and temporal distances (Gumienny et al. 2011). This collaboration can be real-time synchronous—when people are able to work at the same time—or asynchronous, which means people build on what others did by reconstructing and understanding the past interaction and continuing to work on the same content.

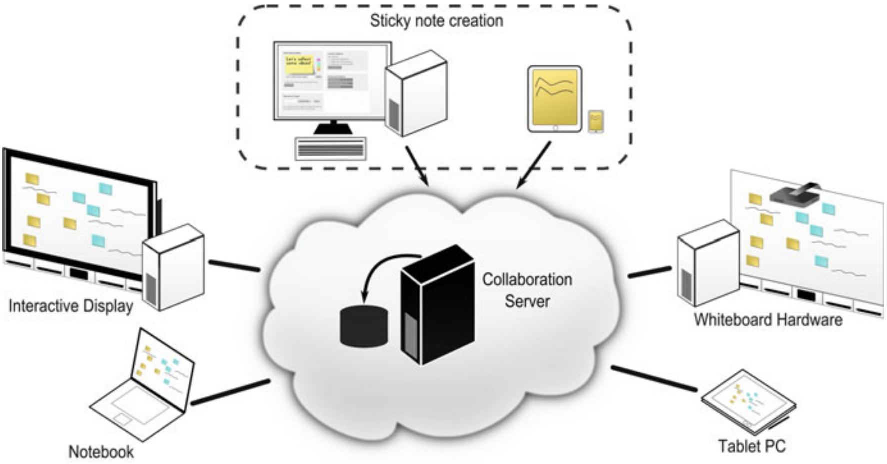


Fig. 1 Outline of Tele-Board system architecture and components' interrelation

The existing system consists of different components (see Fig. 1), each especially designed for fulfilling a specific task:

- Whiteboard Client: Serves as the main user interface by presenting the digital whiteboard drawing surface. It enables people to do the work they are used to on traditional whiteboards by using common metaphors such as sticky notes, hand written sketches, or any other kind of content stuck to the board. Often, a whiteboard client will be started on a large digital whiteboard device, but can also be run on any Java-enabled computer.
- A communication server: Every whiteboard client connects to this server and thereby subscribes to updates of a specific panel. Content is organized in projects and panels and a project can consist of multiple panels. A panel is a whiteboard session, where people are working, and this includes the whole history of their interaction. Changes made on one location are automatically synchronized to all other locations in real time. So people are seeing instantly what people on the other side are doing. For example, while dragging a sticky note, the remote side can follow it moving over the board.
- Tele-Board history: This is a component embedded into the server. Every bit of synchronization communication is captured and immediately stored in a central database. This allows keeping track of the whole interaction process and not only snapshots. Every single operation is stored, which allows applications to resume from any point in time during the past lifetime of the panel, to copy sessions, and run statistics on the usage.
- Mobile devices: Although content can be created directly in the whiteboard client, for an efficient workflow that includes (digital) sticky note pads, we supply people with a variety of applications for mobile devices. Therefore, we developed native applications for different platforms, taking full advantage of all available mobile hardware. By using the apps for iOS or Android devices,

collaborators are able to take pictures, make drawings or just input text. The content is sent to a panel and can be arranged on the whiteboard surface.

3 Broaden System's Availability: Transferring Tele-Board to the Browser

During development of Tele-Board, we had different kinds of hardware in mind. We evaluated this hardware and did prototyping for it. When we started the development of the current whiteboard client, almost all computers had a Java runtime environment installed. Furthermore, Java applications were the most reliable and most common applications running on different platforms.

This changed a lot due to the great success of tablet devices and increased power of mobile phones. The development during the last years changed the way applications are developed and deployed nowadays. New applications need to fulfill the requirement of running on every (mobile) device without having to develop an own version for each platform.

Web browsers are a key component of all mentioned systems. Most computer users are very familiar with these systems. They evolved from applications basically displaying static HTML documents, not well-suited for large scale software development (Mikkonen and Taivalsaari 2008), to software running other software such as web-based games or office systems (Taivalsaari et al. 2011; Anttonen et al. 2011). Many of these media-intensive web applications rely on browser plugins, such as Adobe Flash[1] or Microsoft Silverlight.[2] However, on most mobile platforms, such as Apple iOS or Google Android, those plugins are rarely available.

Compared to most plugin-free browser-based applications, the usability in terms of application start behavior is increasingly cumbersome for Java Webstart applications, such as our Java whiteboard client. When started from the Tele-Board web portal, user action is required in choosing how the application file should be handled by the browser. Due to Java security constraints, users have to explicitly allow the application to be run each time it is started.

In order to overcome these issues, without relying on any browser plugins for the whiteboard client's re-implementation, we build upon the capabilities HTML and JavaScript offer.

[1] http://www.adobe.com/software/flash/about/

[2] http://www.microsoft.com/silverlight/

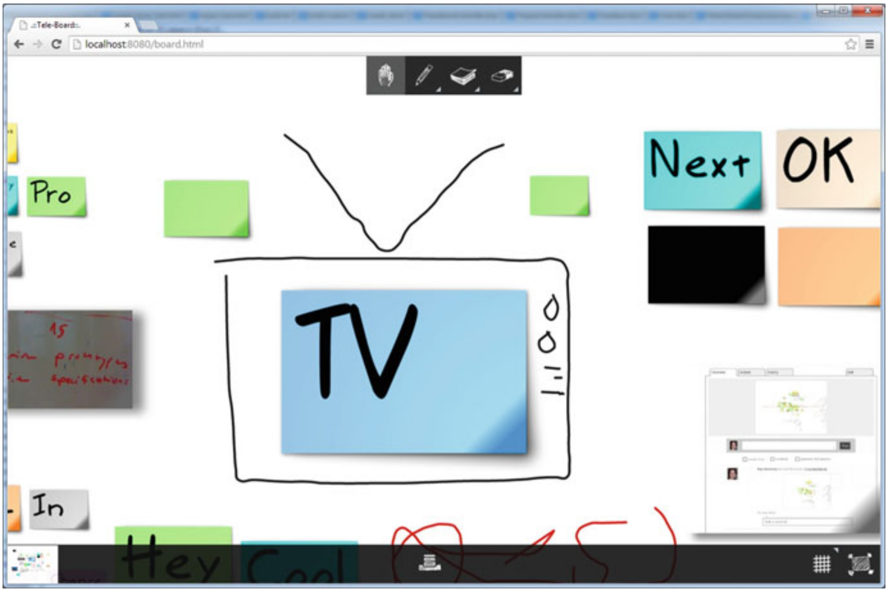


Fig. 2 Web-based whiteboard client in a browser

3.1 Web Browser Implementation Background

We developed a prototype based on HTML5 standards, which resembles the base functionality of the above-mentioned Java Tele-Board whiteboard client. A screenshot of a web browser running our application is shown in Fig. 2. One can see a set of sticky notes and drawings on a virtual whiteboard surface. Zooming and panning functionality is also offered. Furthermore, the use of HTML5 client-server communication ensures synchronization of whiteboard content.

In order to implement a prototype of the Tele-Board whiteboard client that runs natively in a web browser, we identified three main aspects that had to be transferred from the Java application to a browser-based solution:

- Networking—communication with the Tele-Board server component is essential for keeping all connected whiteboard clients synchronized
- Rendering—a virtual whiteboard surface is needed in order to display whiteboard content such as drawings, sticky notes and images
- Threading—networking and rendering have to run in parallel, i.e. a threading mechanism is required to prevent interruption of user interaction when whiteboard content is synchronized

Since the mentioned aspects refer to abstract concepts that are important in many areas they do not apply only to the Tele-Board whiteboard client but to real-time groupware applications in general. HTML5 standard provides mechanisms that apply to those aspects.

Building on the results and recommendations from Gutwin et al. (2011) for browser-based real-time groupware, we use bi-directional *WebSocket*[3] protocol and its corresponding API as networking technology.

The rendering part is implemented using HTML5 *Canvas*.[4] Canvas represents a rectangular area where graphics can be drawn with the help of a JavaScript API in the browser. Compared to other technologies available in modern web browsers, such as *Scalable Vector Graphics*[5] (SVG), the effort for efficient use of Canvas is higher but results in better rendering performance in our implementation.

In order to allow a traceable arranging of sticky notes on the whiteboard surface, a smooth movement animation is required instead of just setting the final position of the sticky notes after moving them. In this way, our implementation requires processing a large number of messages sent to the server, while rendering elements on the whiteboard. The synchronization of whiteboard content and the required network communication is a suitable use case for HTML5 *Web Workers*.[6] Web workers provide a mechanism for parallel task execution within web browsers. This way, networking can be run with minimal interference with the web browser's user interface rendering and user interaction.

Relying on HTML5 standards, modern web browsers, such as Microsoft Internet Explorer 11, Google Chrome and Mozilla Firefox are supported by our implementation.

Implementation details are described in Wenzel et al. (2013).

3.2 Tele-Board in the Web Browser

The whiteboard client's re-implementation results in a system that is fully runnable inside a web browser without requiring any plugins to be installed. The extended system architecture is shown in Fig. 3.

Basically, Tele-Board can be used on any device that is equipped with a web browser. In particular, this includes mobile devices such as mobile phones or tablets (see Fig. 4).

Furthermore, the above-mentioned usability drawbacks, originating from the applied Java technology, were overcome. The web-based client can be opened like any other website by clicking on a link in the web portal. This familiar concept allows seamless usage without requiring further user interaction.

Making Tele-Board accessible by a broader range of hardware adds more flexibility to the remote collaboration process and how users can participate in it more easily. However, this applies only if Tele-Board is already in use and the design team already operates in the digital world. The next section addresses the

[3] https://tools.ietf.org/html/rfc6455

[4] http://www.w3.org/TR/2013/CR-2dcontext-20130806/

[5] http://www.w3.org/Graphics/SVG/

[6] http://www.w3.org/TR/workers

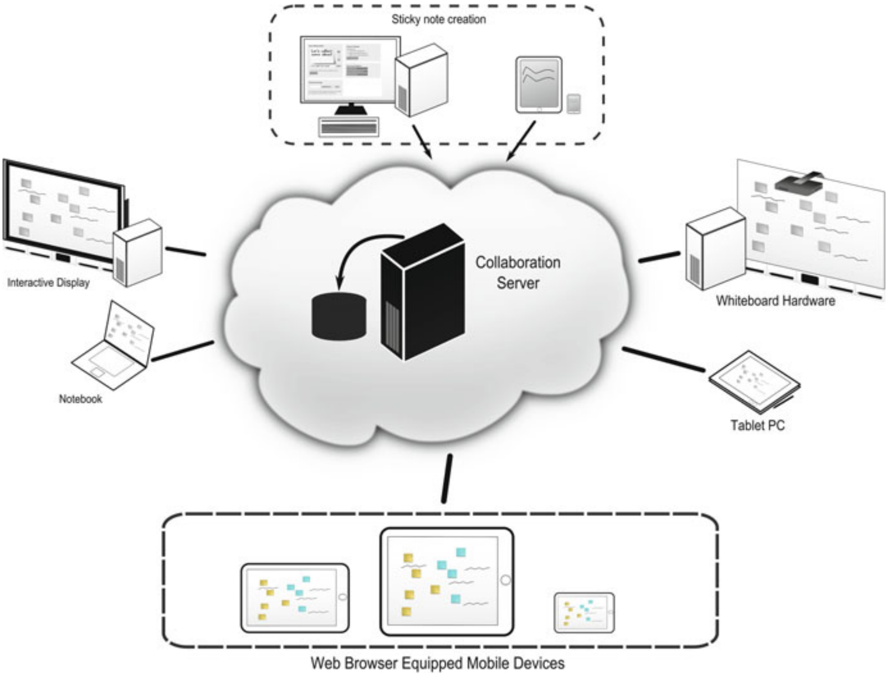


Fig. 3 Extended Tele-Board system architecture. Client-side parts are fully runnable inside a web browser

question of how to ease entry into the digital world of Tele-Board when there are real world artifacts to build upon and these need to be transferred efficiently.

4 Easing System's Application: Bridging the Gap Between Analog and Digital Worlds Through Whiteboard Detection

Tele-Board turns out to be a suitable replacement for regular, analog whiteboards. This includes multiple people interacting with the whiteboard content, ideally simultaneously. Tele-Board does support a wide range of those interaction technologies from IR enabled beamers[7] and eBeam[8] to SmartBoards.[9] While projectors and other temporary additions have proven to be unreliable, e.g. going out of sync

[7] Dell short-throw interactive projector. http://www.dell.com/ed/business/p/dell-s300wi/pd. (November 30, 2013).

[8] eBeam. http://www.e-beam.com/. (November 30, 2013).

[9] SMART Board interactive whiteboards. http://smarttech.com/smartboard. (November 30, 2013).

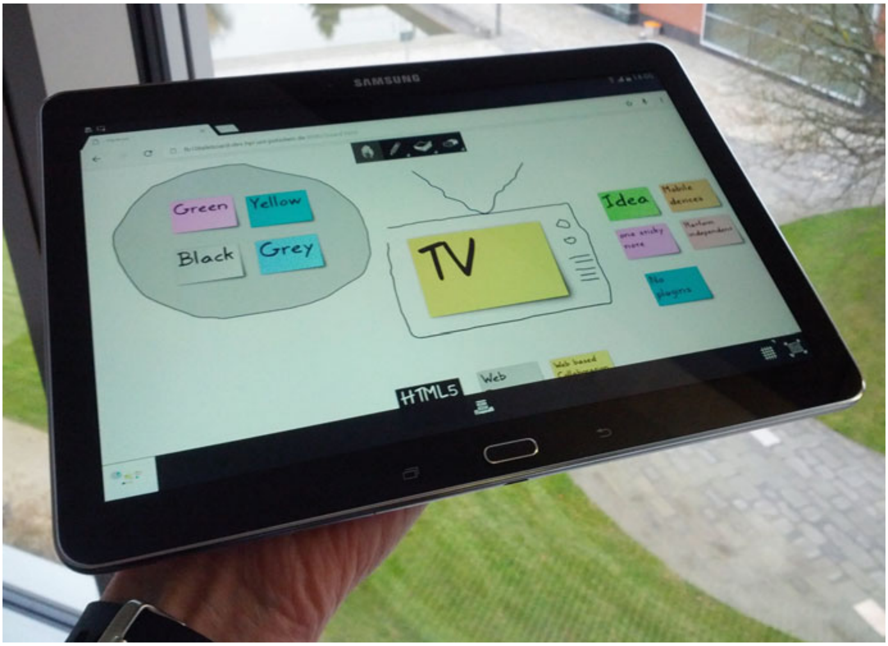


Fig. 4 Web-based whiteboard client in a tablet browser

requiring recalibration every so often, a Smart-Board is quite a different experience. It offers precise input and has special felt tip pens and erasers that mimic the analog experience quite nicely. As amazing as the user experience is, it is equally expensive. With a massive price tag comes the fear of theft, resulting in the fact that such whiteboards are often mounted to the wall eliminating any flexibility in setting up a workplace. Last but not least, substituting the whiteboard for a PC is not great for collaboration just as any natural feeling is lost when drawing with a mouse. At least for now, mobile devices only support partial functionality.

Tele-Board needs technology to work and the more expensive, the better the experience. However, this statement ignores the possibility of technology simply failing and the time it takes to set it up.

Research suggests that for some switching screens could result in a plethora of whiteboards around the workplace (LoBue et al. 2011), others naturally believe differently, leaving the numbers game as its own a disputed problem. However, it becomes clearer when we look at situations where instant creativity is required. A fast brainstorming to gather thoughts on an idea calls for spontaneity. “I have an idea. Let’s try this out.” means: turn around, grab a whiteboard and start drawing, trying and refining. It would hardly help the experience if someone responds: “Hey, think about it while I find us a room and power up the system.”

The benefits of instant global collaboration are less significant in a synchronous setting where the main task is to collaborate. In essence we are looking for a tool that is as lightweight and integrated as a regular whiteboard, yet provides all the amazing features for collaboration when needed. At this time, the final tool should not involve redoing the whiteboard in Tele-Board or settling for immutable photos. The aim is to design a system that allows for a seamless transition from an analog whiteboard to a Tele-Board panel. Achieving this would allow teams to utilize regular whiteboards to experience benefits when work has already been started on a traditional whiteboard:

- Because there were not enough digital whiteboard devices available
- Because using a regular whiteboard was more convenient
- Because it didn't really matter and a regular whiteboard was the preferred choice

Such a symbiosis could help

- with the simple use of Tele-Board as a documentation tool (including handwriting recognition (Gericke et al. 2012b) and search)
- document the development of an idea by saving different states of a whiteboard via the history
- start an idea transfer, whereby other teams are able to modify rather than simply view contents

While the idea is of digitizing analog whiteboard artifacts is not new and something that has been tried with photo and video processing; all approaches had major drawbacks, concentrating on single areas too small to stand a chance of providing support for Design Thinking sessions. Image processing techniques came a long way but with handwriting working out well they are often completely oblivious to the other entities like sticky notes. Other systems are able to recognize them with multiple camera setups, which can potentially solve the problem. This solution incorporates the very same limitations that led to tackling the problem in the first place.

Our goal is a solution integrated into the Tele-Board landscape that allows a Design Thinking session to be transformed into a digital Tele-Board panel (see Fig. 5). This solution should allow for minimal to no loss and thus satisfy requirements of Design Thinkers, who expect a solution to be easy to use and self-explanatory. A software realizing this set of tasks should:

Fig. 5 General idea: transfer manual notes and scribbles into Tele-Board

- Provide instant gratification with a straightforward user interface that can accomplish the task without requiring hundreds of parameters to be set.
- Require a minimal introduction if any at all. A short video demonstrating the process should be enough to show what can be done.
- Be non-intrusive by not imposing too many restrictions on supported platforms or requiring alterations to the workspace.
- Be reliable and reproducible. It should work without major adjustments and should deliver the same results if run twice thus not appearing random to the user.
- Be able to detect all used entities possible on a regular whiteboard. Most notably it should be able to distinguish text and sticky notes and should transform a sticky note into its digital representation with the same colors and editable text.

4.1 Process and System Landscape Overview

This section will give an overview of the process used to achieve the analog/digital transformation. Besides giving a rough idea it will specify this process in more detail. With this process in mind, the following section will introduce the final application "in action." Via a system walkthrough the theoretical process will be linked to the workflow as it is outlined for the user of the proposed solution.

With Tele-Board being a digital tool, all interaction and input is digital as well. Our application introduces adding whiteboard content that does not originate in the digital world. This approach, allowing the transition of an analog whiteboard into the Tele-Board landscape, means it will finally be possible to directly proceed with a digital setup.

The main work is focused on how we can enable classical whiteboard sessions with global digital collaboration. As stated before, the question relies on a synthesis of classical whiteboards with digital collaboration tools. While with fully digital workspaces Tele-Board already achieves this goal, mixed workspaces, especially in asynchronously working teams, are still dependent on the answer as to how might we bridge the gap between analog and digital whiteboards (Fig. 6).

Any acceptable solution has to be non-intrusive in nature. This means that overhead for the Design Thinking teams must be kept to a minimum. Capturing the most accurate information as well as a detailed timeline would be possible with a live capturing system. However this would include a lot of technology that has to work properly and be set up in the beginning. This effort would cause a barrier hindering widespread use. The alternative is taking photos of whiteboards. This is not time consuming especially with the common practice of project documentation in mind. Therefore taking pictures at important junctures is already part of the workflow and the process just has to tweak this step in order to get optimal results. A major disadvantage of this kind of setup is a missing detailed timeline. The photo represents merely the final thought product. Nevertheless, photo documentation is acceptable for the following reasons:

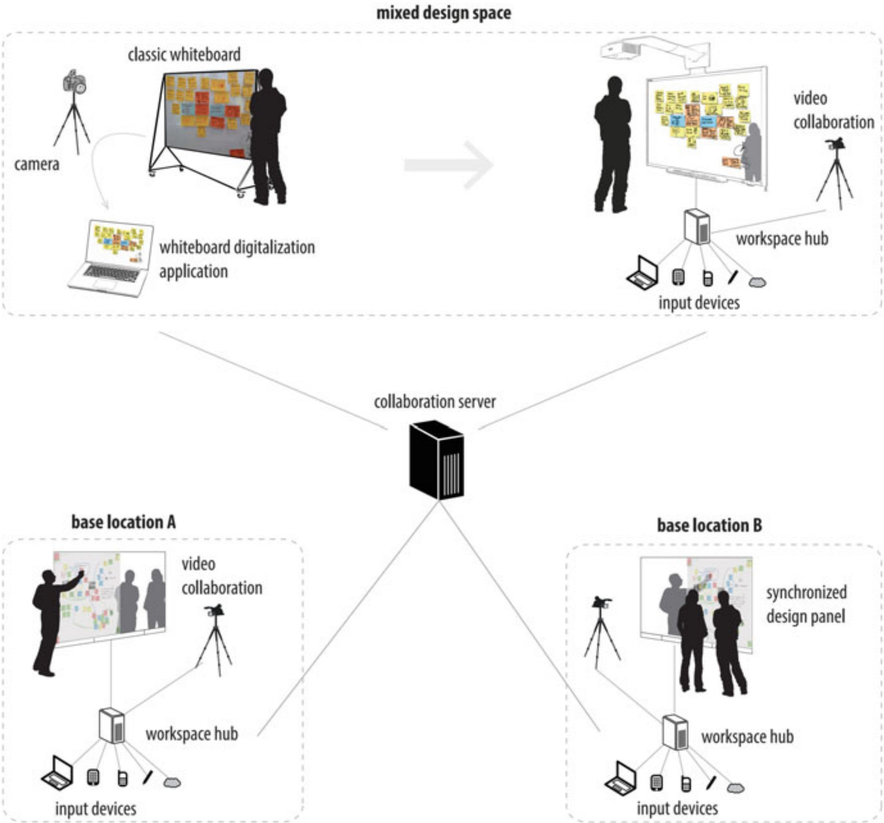


Fig. 6 System landscape and how it broadens the scope of regular Tele-Board setups

1. The simple transfer of an analog co-located discussion into a synchronous distributed collaboration does not necessarily require a history. A seamless transition allows teams to start their work in any form they like with the ability to later switch to global collaboration.
2. Taking pictures is already part of the workflow. Thus the only required tool is already at hand. It is the least intrusive way of allowing this transition to happen. Most other solutions would require stationary and complex setups that run contrary to the very basic motivation for this transition.
3. It is possible to keep track of major changes over time. Again, for documentation purposes, not only the absolute final result is photographed but also important milestones on the way. This photo set does not reproduce the fine detail of the digital whiteboard, rather it represents events and states the team considers important. These photos can be transformed into different Tele-Board panels as well as multiple states on one panel providing a temporal chain of changes.

The following will introduce the process used for the transition of a whiteboard photo into a Tele-Board panel. Based on the rationale described in the previous paragraph, it centers on whiteboard pictures.

In essence, the photo of a whiteboard to digitize will be handed over to an application which, by using image enhancement and image processing methods, identifies the whiteboard, notes, drawings and any other content. After this point the content is imported to the Tele-Board portal to be used as any other regular Tele-Board panel.

Naturally—as shown in Fig. 7—the algorithm consists of several steps necessary to achieving the digitalization of a whiteboard. An aspect, not directly part of the process, is the camera used for the whiteboard image. While there are no direct requirements for a specific camera, the image quality has a direct effect on the performance and reliability of the algorithms. Thereby, without requiring high-end equipment, any chance of increasing image quality will likely be rewarded by more detailed and complete end results.

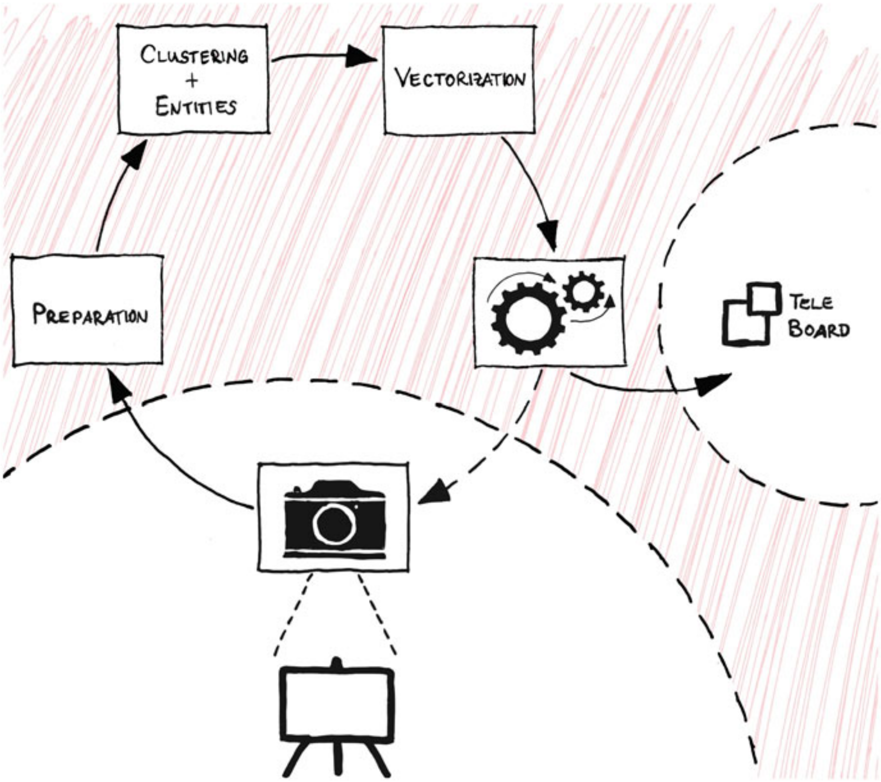


Fig. 7 Depicting the steps taken, starting with a whiteboard photo, to achieve a digital version of the same information content

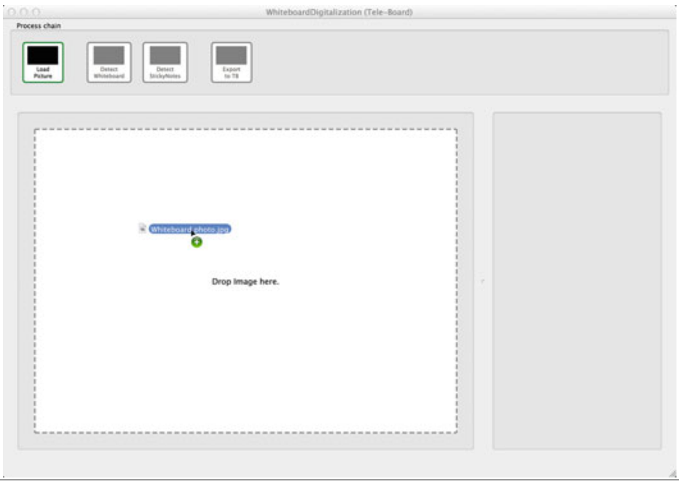
Subsequently, the process is able to start deducing actual information. This process has two parts. First, the image will be analyzed to understand what elements can be found and what type they are, i.e. handwriting on sticky note. In a second step, a particular element will be transformed into a digital, scalable element. At this stage the algorithms have deduced how many sticky notes there are, what color they are and where they are located on the whiteboard. Moreover, the writing and drawing on the board, as well as the sticky notes are available in vector format.

An export step will provide the necessary transformations and final processing steps. It will calculate the location of a sticky note based on its relative position on the original whiteboard, or upload an image used in the whiteboard in order for it to be properly displayed inside a Tele-Board panel. After the export step is completed the digital representation is stored on the Tele-Board server as a regular panel. The representation is accessible via the Tele-Board portal and Tele-Board whiteboard client. The conversion is complete.

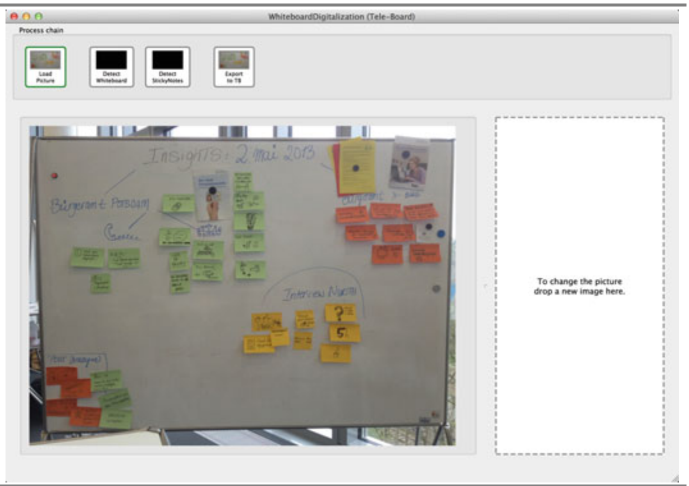
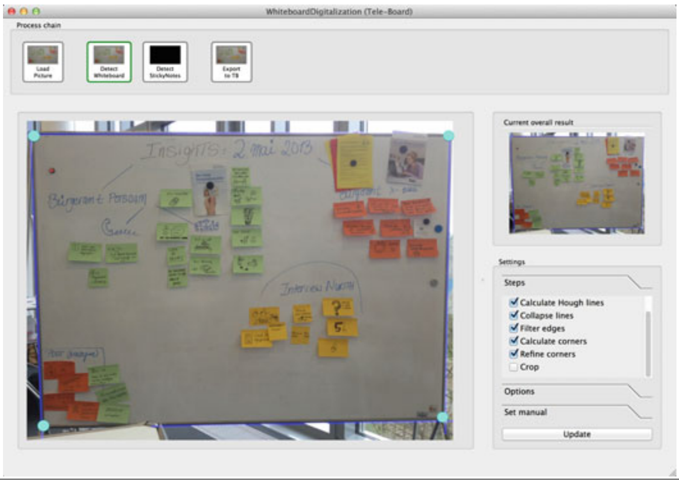
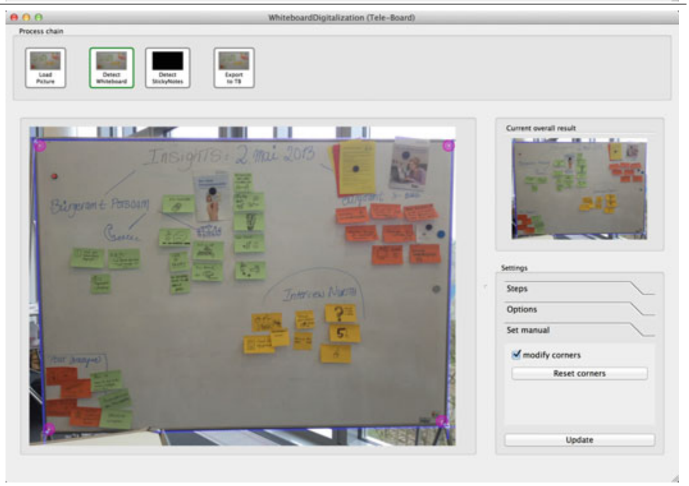
The local, static whiteboard image, taken by a camera, has become a virtual, shared whiteboard surface, where drawings, sticky notes and images can be handled individually. Handling includes moving, scaling or erasing them in distributed environments.

4.2 *Real Life Application Walkthrough*

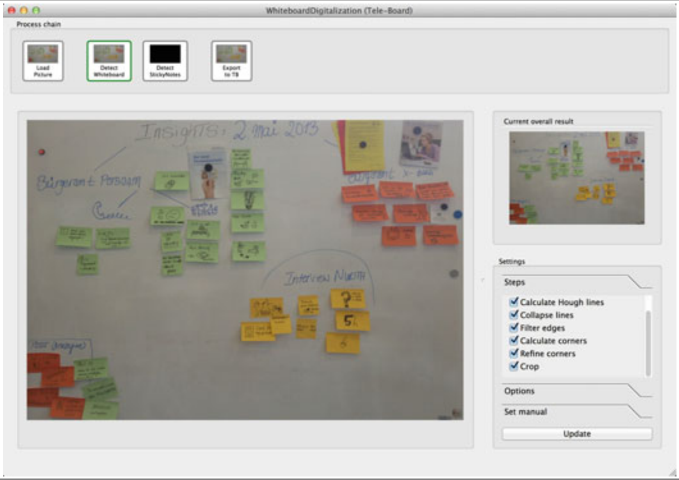
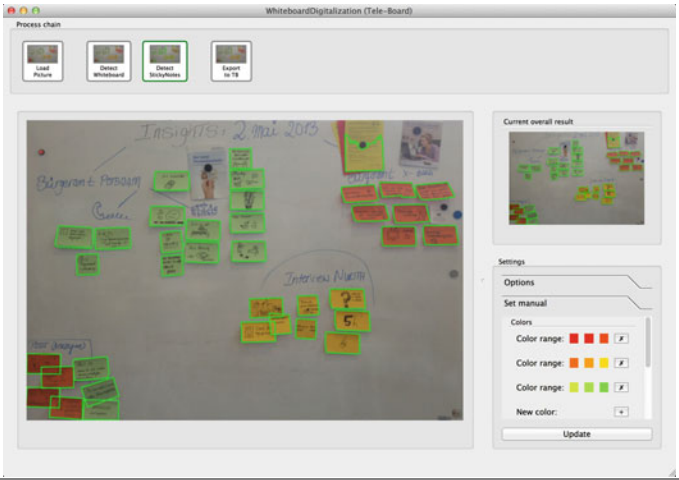
The developed application is presented in the following system walkthrough in the setting of regular intended use. For process step annotations see Fig. 7.

Description	Illustration
Start The application starts with a minimal user interface introducing the process steps and main stage prominently displaying a drop area. This is the easiest way of adding a document to a non-full screen application.	

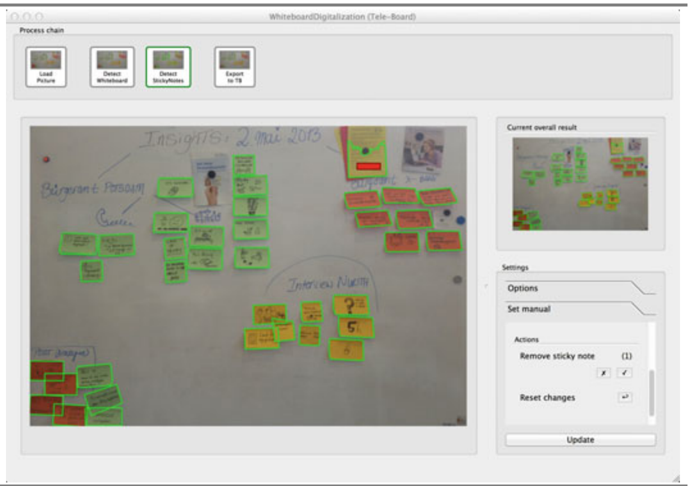
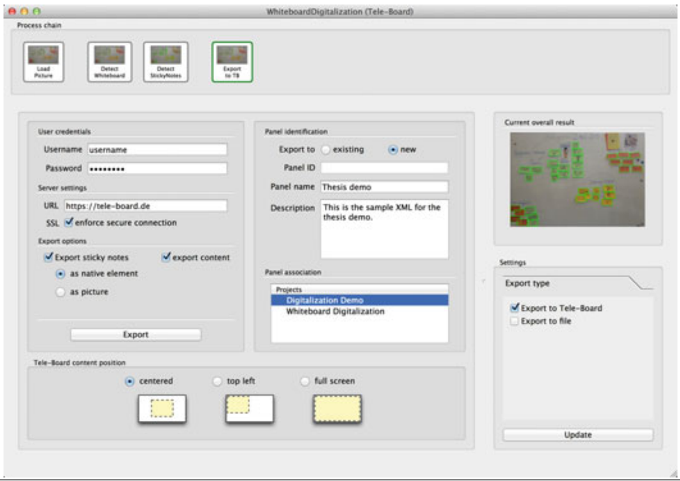
(continued)

Description	Illustration
[start] The photo chosen will be displayed. The right side morphed into a new drop zone. It can be used to start working on a new photo at any time. Additionally the other process steps are unlocked now and the user can proceed.	
[preparation] As the image shows more than just the whiteboard itself, Detect Whiteboard is used. The result is presented in the center with the identified corners and option for manipulation on the right.	
[preparation] Unsatisfactory results can be corrected. The manual options pane on the right allows making the detected corners moveable by mouse. This mechanism provides an automated suggestion suitable to most situations combined with the power of interference.	

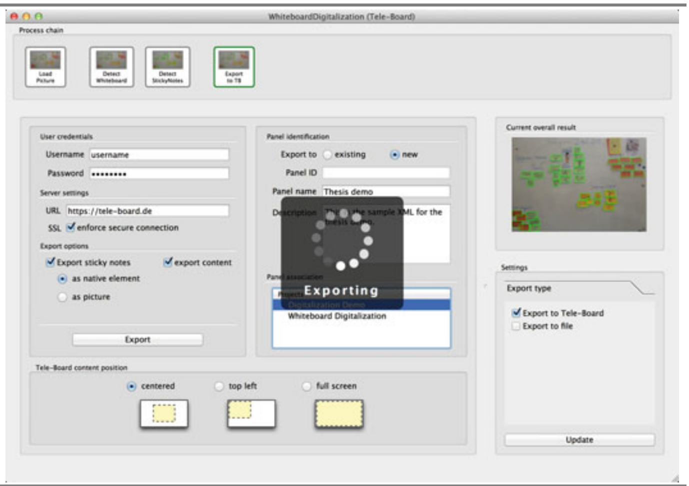
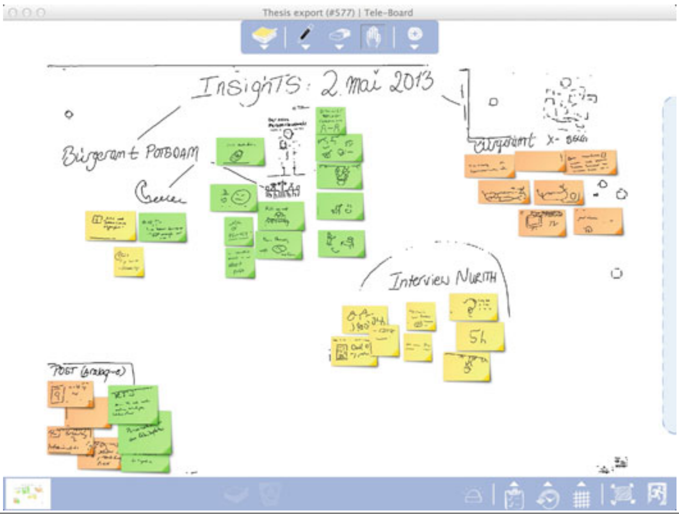
(continued)

<table>
<tr><th>Description</th><th>Illustration</th></tr>
<tr><td>[preparation] Once the whiteboard is marked it is time to extract it. Cropping is a central step as it is the last preparation step and produces the image that all the other algorithms rely upon.</td><td>
</td></tr>
<tr><td>[entities] Next in line is entity detection or more specifically sticky note detection. The image of the previous step is used to search for any sign of sticky notes. If so, the contours will be superimposed for the user to check the result.</td><td>
</td></tr>
<tr><td colspan="2">Complex clusters will automatically be separated. This feature is less reliable than others and is optional via a user interface switch. If the algorithm missed a certain sticky note color or mistakenly identified one, the manual option pane provides tools to remedy both situations. Setting changes will cause reprocessing.</td></tr>
</table>

(continued)

<table>
<tr><th>Description</th><th>Illustration</th></tr>
<tr><td>[entities] Additionally, wrongly identified sticky notes can be removed. A marker will be placed in its position reminding the user why this area was not recognized. This is a helpful tool to remove other entities, e.g. a document as shown on the right.</td><td>
</td></tr>
<tr><td>[export] The export form allows specifying all options needed for the final import in the Tele-Board system. User credentials and system URL are obvious choices. As the whiteboard becomes a Tele-Board panel, some information about the new panel has to be addressed.</td><td>
</td></tr>
<tr><td colspan="2">The bottom pane allows the user to specify how the content should be scaled and displayed. Internally it collects all the data, packages it and handles the communication with the Tele-Board portal server.</td></tr>
</table>

(continued)

Description	Illustration
[vectorization] Content detection is where the whiteboard drawings become vector information that can later be manipulated in the whiteboard client, e.g. be erased. It is a silent component that is called as part of the export process, hence no user interface for the user.	
[tele-board] Finally, the whiteboard photo has been transferred into a Tele-Board whiteboard and can be used by the whole team globally with the whiteboard client. If another whiteboard needs to be digitalized, the user could start from the beginning using the drop side panel.	

5 Conclusion

Design Thinking methodology is often characterized by a flexible working mode not only as it applies to the spatial environment but also to the utilized tools. Tele-Board is a remote collaboration software tool supporting Design Thinking working modes, and traditionally implemented in co-located setups. However, a transition from analog, co-located to digital Tele-Board supporting remote collaboration settings was hardly possible. The decision to use our system had to be made in the beginning of the working process.

We integrated a new tool in Tele-Board's system landscape, closing the media gap existing as long as digital whiteboard tools have been around. It renders assistance to all those who want a digital tool as universal and flexible as a traditional whiteboard, yet offering all the benefits of the digital world. The application starts with a whiteboard photo, a step all too familiar to Design Thinking teams. By applying advanced image processing algorithms, a digital clone of the former analog whiteboard is created, enabling classical whiteboard sessions with global digital collaboration.

A way was found to overcome obstacles in digital whiteboard use by removing the bar for switching to digital whiteboards from a regular whiteboard. This formerly optional technology now becomes yet another tool for Design Thinking, readily accessible whenever needed.

Arriving at the digital world, the software tool becomes an integral part of the working process in a design team. As a team player, Tele-Board has to adapt, in terms of supported hardware devices used for interacting with the system. Mobile devices such as mobile phones and tablets are more and more important in today's working environments. The previous version of Tele-Board's whiteboard client component was based on Java technology, which is not available on mobile platforms. Furthermore, starting this application became increasingly cumbersome. Hence, we re-implemented this component on the basis of modern web technologies allowing its seamless usage from within web browsers on desktop as well as mobile systems while avoiding Java technology constraints. This way, it allows a more flexible participation in remote collaboration settings.

By consequently following the web-based approach, future development will focus on server side whiteboard detection. The goal is to integrate this component as a web-based application making Tele-Board an integrated system running in a browser and therefore accessible by the maximum number of available devices.

References

Anttonen M, Salminen A, Mikkonen T, Taivalsaari A (2011) Transforming the web into a real application platform: new technologies, emerging trends and missing pieces. In: Proceedings of the 2011 ACM symposium on applied computing—SAC '11, p 800

Gericke L, Gumienny R, Meinel C (2012a) Tele-board: follow the traces of your design process history. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer, Berlin, pp 15–29

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 2012)

Gumienny R, Meinel C, Gericke L, Quasthoff M, LoBue P, Willems C (2011) Tele-board: enabling efficient collaboration in digital design spaces across time and distance. In: Plattner H, Meinel C, Leifer L (eds) Design thinking, understand—improve—apply. Springer, Berlin, pp 147–164

Gumienny R, Gericke L, Wenzel M, Meinel C (2012a) Tele-board in use: applying a digital whiteboard system in different situations and setups. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer, Berlin, pp 109–125

Gumienny R, Hampel S, Gericke L, Wenzel M, Meinel C (2012b) Transferring traditional design work to the digital world: does it work? In: Proceedings of the 2012 design research society international conference (DRS 2012)

Gutwin CA, Lippold M, Graham TCN (2011) Real-time groupware in the browser: testing the performance of web-based networking. In: Proceedings of the ACM 2011 conference on computer supported cooperative work, pp 167–176

LoBue P, Gumienny R, Meinel C (2011) Simulating additional area on tele-board's large shared display. In: HCI international 2011—posters' extended abstracts, vol 173, pp 519–523

Mikkonen T, Taivalsaari A (2008) Web applications—Spaghetti code for the 21st century. In: 2008 Sixth international conference on software engineering research, management and applications, pp 319–328

Taivalsaari A, Mikkonen T, Anttonen M, Salminen A (2011) The death of binary software: end user software moves to the web. In: 2011 Ninth international conference on creating, connecting and collaborating through computing, No 978, pp 17–23

Wenzel M, Gericke L, Gumienny R, Meinel C (2013) Towards cross-platform collaboration—transferring real-time groupware to the browser. In: Proceedings of the 2013 I.E. 17th International conference on Computer Supported Cooperative Work in Design (CSCWD), pp 49–54

Diagnostics for Design Thinking Teams

Neeraj Sonalkar, Ade Mabogunje, Gina Pai, Aparna Krishnan, and Bernard Roth

Abstract Multidisciplinary teamwork is a key requirement in the design thinking approach to innovation. The tools currently available for effective team coaching are limited to heuristics derived from either experienced design thinking professionals or clinical psychology practitioners. Our research aims to improve this current situation by providing design thinking managers, coaches and instructors a scientifically validated tool for augmenting design team performance. We present the development of a software tool called the IDN Tool based on the Interaction Dynamics Notation to analyze team interactions and diagnose patterns of behavior that influence design outcomes. We demonstrate the use of the IDN Tool through analysis of the interaction behaviors of seven design teams engaged in a concept generation activity, which were independently rated by a two-person Jury using the criteria of utility and novelty. Through the analysis we were able to visually isolate the interaction behaviors that had a high positive or negative correlation with the levels of novelty and utility of concepts judged a priori. With further work, this has the potential of improving in-process design team performance with a positive influence on design outcomes.

1 Introduction

Design Thinking as an approach to the development of new products or services emphasizes three key elements—user empathy,[1] iterative prototyping,[2] and multidisciplinary teamwork. Individuals from different disciplines, departments, and

[1] User empathy and perspective-taking.

[2] Ideation, prototyping, and testing.

N. Sonalkar (✉) • A. Mabogunje • G. Pai • A. Krishnan
Center for Design Research, Stanford University, Building 560, 424 Panama Mall, Stanford, CA 94305, USA
e-mail: sonalkar@stanford.edu

B. Roth
Hasso Plattner Institute of Design, Stanford University, Stanford, CA, USA

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_4

different stakeholder groups participating in a design thinking project work in teams to understand user needs, develop product concepts, prototype and test concepts with users to come to a product outcome that meets user needs. Teamwork thus underlies all the activities that comprise design thinking projects. However, this crucial element of teamwork is neither well understood nor appropriately supported in practice. In industry and academia, design managers and instructors often put individuals into teams and take them through the various activities of design thinking with only little understanding about characteristics that make a design team effective. The tools for teamwork coaching are limited to heuristics derived from either experienced design thinking professionals or clinical psychology practitioners.

On the other hand, design-thinking research has investigated teamwork since 1990s (Tang and Leifer 1991; Cross et al. 1996; Valkenburg 2000). But this research has not had any significant impact on design thinking practice. Even now, recent research findings about affect expression (Jung 2011), team composition (Schar 2011; Kress 2012), and idea generation in teams (Edelman 2011) do not influence the practice of design thinking in industry or academia. This situation reflects a knowing-doing gap (Pfeffer and Sutton 2013) between research and practice.

How do we overcome the knowing-doing gap between research and practice? We propose that ***the understanding of teamwork emerging from research studies needs to be embedded in diagnostic instruments that are useful to practitioners***. The development of instrumentation has been a key factor in the development of medicine as a science based practice. Instrumentation for diagnosis and intervention has enabled scientific discoveries to be available at hand for medical practitioners to use in the messy real-world situations that confront them. In a similar vein, we propose that design research needs to develop instrumentation for diagnosis and intervention that enables research discoveries to be amenable for practical application. In this chapter, we present the development of a diagnostic system for design teams that integrates research findings about team behavior in a visual form that is amenable for application to practice.

2 Research Questions

The questions guiding our research in the development of a visual diagnostic system for design thinking teams were as follows.

1. What is the spectrum (or atlas) of discernable and significant interaction patterns that occur in design thinking teams?
2. How do these interaction patterns influence design outcomes?

A diagnostic instruments needs to not only detect patterns but also indicate what the patterns signify in terms of design outcomes. Question 1 refers to the identification of interaction patterns in design teams. Question 2 refers to the significance of the identified patterns in terms of influence on design outcomes. The following section describes the method followed to investigate these questions.

3 Method

We adopted an engineering design approach to develop and test a visual diagnostic instrument. Our prior research (Sonalkar 2012; Sonalkar et al. 2013) had led to the development and validation of the Interaction Dynamics Notation. The notation had shown that through effective use of a visual symbol system, it was possible to integrate previous research on team dynamics into a single analysis system while retaining the moment-to-moment temporality of team interaction. The Interaction Dynamics Notation was chosen as the foundation on which to build a visual diagnostic instrument for design thinking teams. An overview of the Interaction Dynamics Notation is given in Sect. 4.

Once the Interaction Dynamics Notation was chosen to build the visual diagnostic instrument, further design requirements for such an instrument were identified and multiple prototypes were built to satisfy these requirements. These were iteratively improved through testing on a databank of videos of teams engaged in concept generation activity. This resulted in the development of the IDN Tool. The development of the IDN Tool is discussed further in Sect. 5. Section 6 describes the specifications of the IDN Tool.

The capability of the IDN Tool as a diagnostic instrument was tested with a dataset consisting of concept generation interactions of seven teams. The concepts generated by the teams were analyzed in terms of their novelty and utility to generate outcome measures. The patterns identified through the IDN Tool were correlated to the outcome measures to identify the interaction patterns that could have an influence on design outcomes. Sects. 7 and 8 describe the application of the IDN Tool to concept generation interactions and the detection of interaction patterns correlating with design outcomes.

4 Interaction Dynamics Notation: An Overview

Interaction Dynamics Notation creates a descriptive visual model of team interaction by interpreting and assigning symbols to observable speaker expressions (verbal and nonverbal). The assignment of symbols is conducted based not on what the expression is from the point of view of the person making it, but on what the expression is taken to be and responded to by others in the team. So in effect we are modeling a series of speaker responses rather than a series of speaker expressions. Thus, the Interaction Dynamics Notation is a visual model of an unfolding interaction. Figure 1 shows the Interaction Dynamics Notation of a brief design conversation.

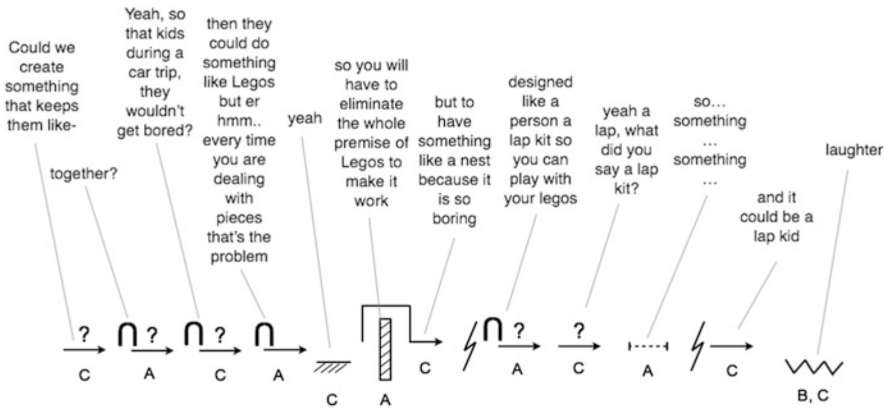


Fig. 1 A conversation between three designers A, B and C is visualized using the Interaction Dynamics Notation

Table 1 gives a detailed explanation of each symbol used in the visual notation. For further information about the development of the Interaction Dynamics Notation, please refer to Sonalkar et al. (2013).

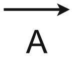
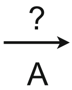
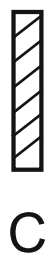
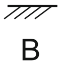
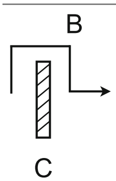
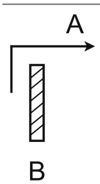
5 Development of the IDN Tool

The initial research on the Interaction Dynamics Notation depended on paper and pen based manual analysis of video. This was a tedious and time-consuming process. In order to use the Interaction Dynamics Notation in a visual diagnostic instrument viable for real-world use, it was necessary to develop a software tool that would accelerate the use of the notation. This software tool would then form the core element of a visual diagnostic system for design thinking teams. The following design requirements were identified for developing this software tool, which later became known as the IDN Tool.

5.1 Functional Requirements

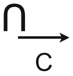
1. The IDN Tool must work with video data as well as conversations happening in real-time.
2. The IDN Tool must give a visual output in terms of the Interaction Dynamics Notation that is easy to understand.
3. The IDN Tool should include sequential analysis methods in order to accelerate the analysis of team interaction patterns.
4. The IDN Tool must work on both Windows and OSX platforms.

Table 1 The visual symbol set for the Interaction Dynamics Notation

Symbols	Name	Description	Example
	Move	A 'move' indicates that a speaker has made an expression that moves the interaction forward in a given direction.	A: I need to buy Legos (at) home. Think about how therapeutic it would be.
	Question	A question indicates an expression that elicits a move. A question projects onto the next response and constrains the content of that response because the next response needs to answer the question.	A: Where should we start?
	Silence	Silence is a state in the conversation when none of the participants speak as they are engaged in other individual level activities. Silence has been included in the notation as a number of design conversations are an interplay of both group conversation and individual activity.	
	Block	Block indicates an obstruction to the content of the previous move. For a block to be felt, the coder needs to feel that the response in some ways obstructed the flow that was established by prior moves.	B: Maybe have something which looks like a computer but you can just type your name or do a simple math, a calculator in the shape of a computer kind of. C: Er, but I don't know, I mean, considering the age segment we are targeting 3–7 years.
	Support for move	Support-for-move indicates that the speaker understands and/or agrees with the previous move.	C: Safe and entertaining (bending forward to write). B: Safe and entertaining, yes.
	Support for block	Support indicates an acceptance of a block by another person.	A: But that's also, I think that's already done. C: Yeah, its already there. B: Ok.
	Overcoming	Overcoming a block indicates that though a block was placed in front of a move, a speaker was able to overcome the block and persist on course of the original move.	C: Er, but I don't know, I mean, considering the age segment we are targeting 3–7 years. B: So 7 years they go to school, they would learn A, B, C right?
	Deflection	When a speaker blocks a previous speaker's move, that speaker or another can deflect the block with a move that presents an alternative direction for the interaction.	B: So when you say we need to divide the age-group, but you cannot have like 3, 4, 5. A: No, no of course not, but I mean you might have a few different (concepts).

(continued)

Table 1 (continued)

Symbols	Name	Description	Example
	Interruption	An interruption is indicative of a speaker being interrupted by another speaker or at times by himself.	B: Should we start generating some concepts now? A: Yeah (interrupted by X). X: 10 min are gone.
	Yes and	A move is considered to be a 'Yes and' to the previous move if it accepts the content of the previous move and adds on to it.	A: What about… if we made a toy that incorporates girls and boys. Its like a house that has a car with it kind of like enables the guys to play with the girls? C: I think that's a good point to have some sort of a educational point in it.
	Deviation	Deviation indicates a move that changes the direction of the conversation from the one implied by the previous moves.	C: But we need to remember it. C: This is not the buildable room (deviating from previous topic).
	Humor	Humor indicates instances of shared laughter in teams.	A: I don't know I probably would have swallowed but (All of them laugh).

5.2 User Interface Requirements

1. The user interface must be easy to use.
2. The user interface must accelerate the rate of video coding as compared to manual coding.

With these requirements in mind, the IDN Tool was developed through iterative prototyping with a set of test video data available from previous research studies (Sonalkar 2012). The following section describes the specification of the IDN Tool.

6 IDN Tool Specifications

6.1 Functional Specifications

The IDN Tool software enables a user, an IDN analyst or researcher to import video data, code the video data so that specific visual symbols are assigned to specific speaker responses, and then output this assignment of visual symbols as a visual representation. The IDN Tool also incorporates a sequential analysis functionality that includes finding patterns of sequential symbol assignments, and conducting a Markov analysis of the probability of one response following another. Figure 2 describes the basic functionality of the IDN Tool.

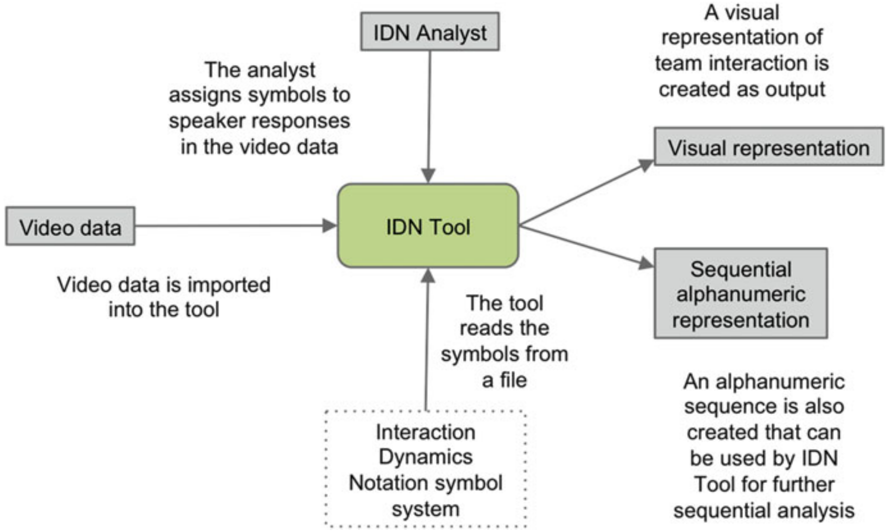


Fig. 2 Overview of the functionality of the IDN Tool

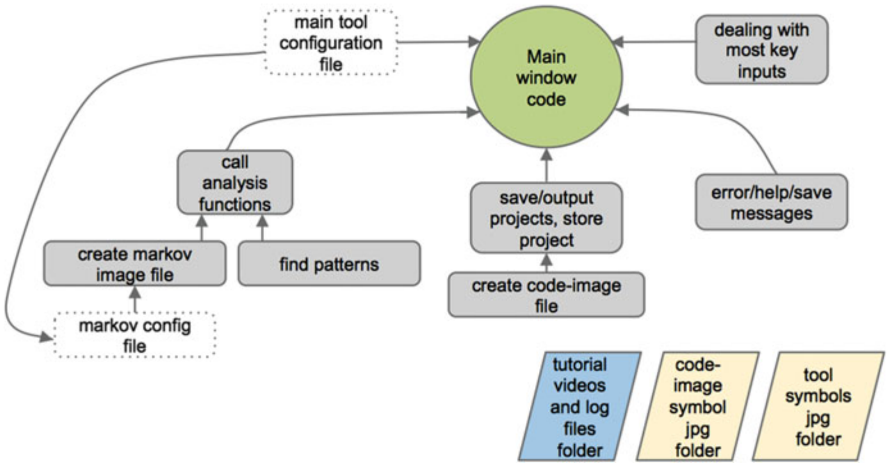


Fig. 3 Functional modules of the IDN Tool

This functionality is achieved through the various functional modules that are part of the IDN Tool as described in Fig. 3, and the file structure described in Fig. 4.

The IDN Tool is coded in Python language on a Linux platform. The decision of using Python on Linux was made in order to create a software code that could then be easily ported to Windows or mac OSX platforms.

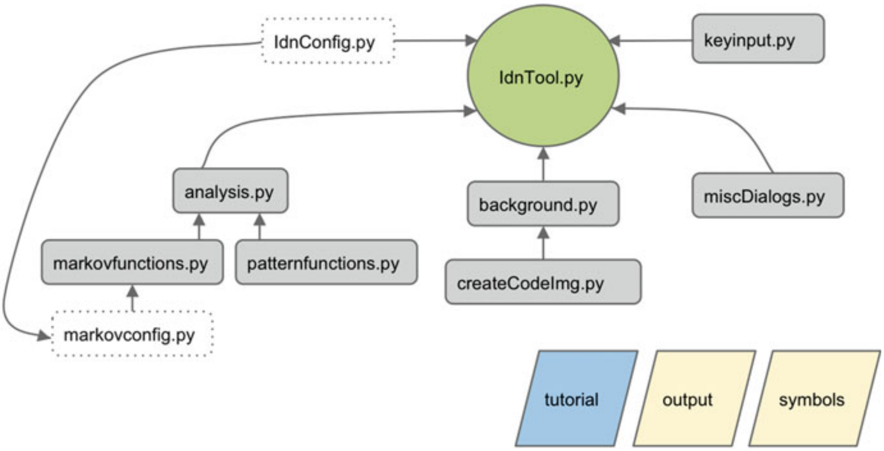


Fig. 4 The corresponding files architecture that forms the IDN code

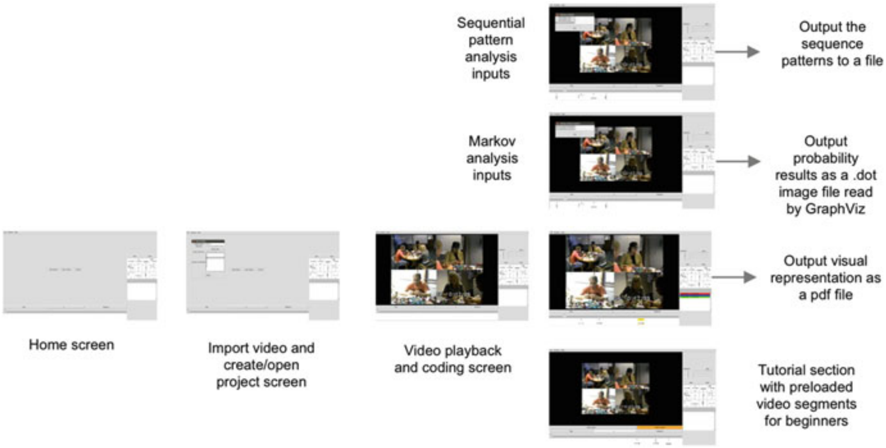


Fig. 5 Overview of the user interface screens of the IDN Tool

6.2 User Interface Specifications

Figure 5 describes the various screens that are part of the IDN Tool user interface.

The main video coding screen is shown in Fig. 6. The video is displayed prominently in the middle. The lower display bar shows the symbols that are being coded. The right-hand panel shows the speaker label assignments, and the hot key assignments for each symbol. It also includes a toggle button for coding start and stop of topic segments.

Fig. 6 The interface elements of the video coding screen of the IDN Tool

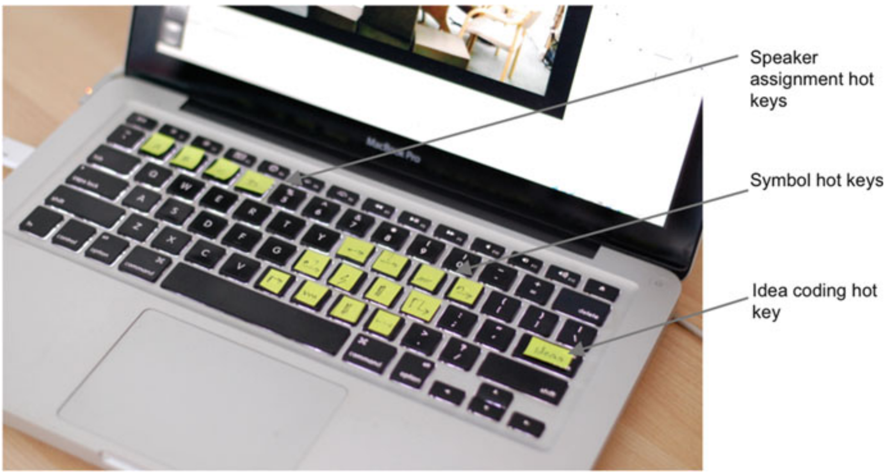


Fig. 7 Hot key labels on the keyboard to accelerate video coding

Once the coding is completed, the file can be saved and the visual output representation and the alphanumeric representation can be exported from the main menu bar.

In order to accelerate the coding of video, the IDN Tool can be operated through a set of hot keys. The keyboard is modified by overlaying the IDN symbols on the hot key buttons as shown in Fig. 7.

7 Analyzing Design Team Interactions with IDN Tool

Two team interaction analysis studies were conducted using the IDN Tool. The first was a comparative study of the IDN Tool with manual method of coding the Interaction Dynamics Notation. Four concept generation team sessions each of 40 min were analyzed using the IDN Tool as well as by using paper and pen. The comparison showed that the IDN Tool was 33 % more effective for researchers to code video data as compared to paper-based coding.

The second study was designed to test the version 3 of the IDN Tool with a data set that had measurable design outcome parameters. This study is described in detail in this section.

7.1 Concept Generation Study

Seven teams of three to four participants were given a concept generation task based on a real-world challenge. The participants were chosen from graduate students at Stanford University who had previous exposure to design thinking. The teams were invited in to the Design Observatory (Carrizosa et al. 2002; Törlind et al. 2009), which is pre-configured to record multiple video streams for design activity analysis. The teams were given the design brief that asked them to generate two concepts for the challenge of lifting water from below 50 ft. underground for small holding farmers in Myanmar. The design brief was designed to stimulate conversation and hence it explicitly asked the teams to generate a best-fit concept that would be technically feasible, and a wild idea or a dark horse concept that was unlikely to be feasible, but could revolutionize irrigation if it could work. The teams were given 60 min for concept generation and a further 20 min to sketch their deliverables and submit them to the research team. Figure 8 shows the team interaction setting and Fig. 9 shows a sample concept sketch arising from such team interaction.

The videos of the seven design teams were imported into the IDN Tool and analyzed to create visual representations of their concept generation interaction. The occurrence of ideas during interaction was depicted by highlighting the symbols corresponding to the speaker turns in which ideas were expressed. Figure 10 shows a sample visual representation output from the IDN Tool.

The use of IDN Tool for analyzing concept generation interactions of design teams demonstrated its use as a video coding tool. The IDN Tool generated a visual output, as well as an alphanumeric output of concept generation interactions for seven teams.

The IDN Tool used the alphanumeric output to detect sequences of symbols that occurred more than three times in the concept generation session. Some examples

Fig. 8 Four camera video stream of team design interaction

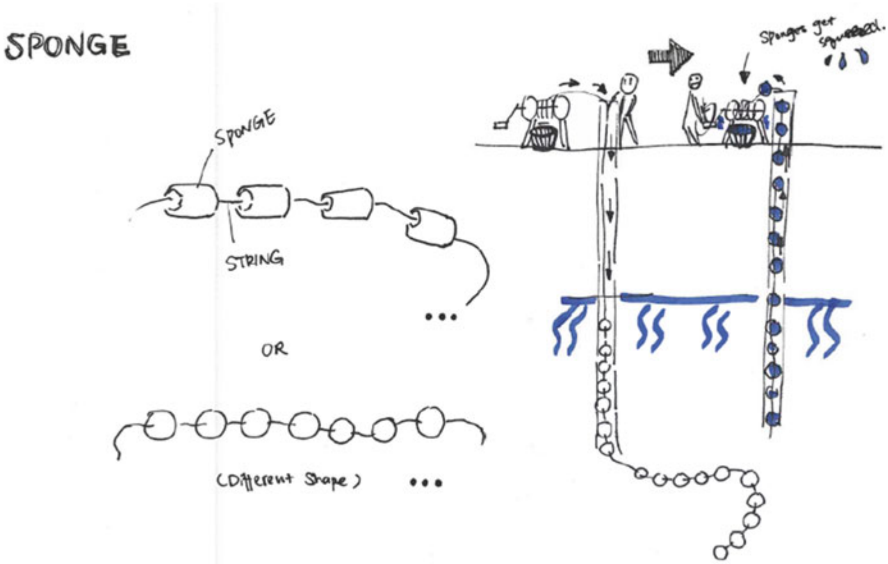


Fig. 9 A concept sketch developed by one of the design teams in the study

of sequences identified in Team 1 are mmqm (12 times), mshm (14 times), and mhms (eight times) where m = move, q = question, h = silence, and s = support. The IDN Tool was able to identify 33 (Team 7) to 142 (Team 2) such sequences in the concept generation data for each of the teams. The identification of such interaction patterns addressed the first research question presented in Sect. 2.

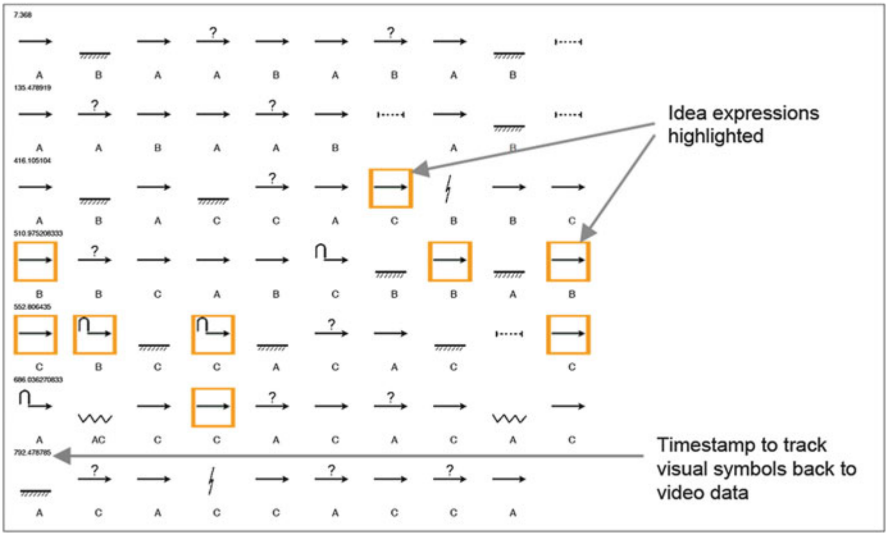


Fig. 10 A sample visual output of team concept generation interactions from the IDN Tool

However, even though these interaction patterns were identified, further analysis was required to address the second research question, and understand which of these patterns correlated with design outcomes.

7.2 *Expert Assessment*

In order to correlate interaction patterns with design outcomes, the concepts generated by the seven design teams were analyzed to obtain outcome measures. We used the ideation effectiveness metric proposed by Shah et al. (2003) to evaluate the concepts proposed by teams on their novelty and utility. The 14 concepts, two for each team were rated by two mechanical engineering experts on parameters pertaining to utility and novelty. Utility parameters included technical feasibility, satisfaction of requirements, manufacturability, serviceability and affordability. Novelty parameters included novelty of mechanism and novelty of human-machine interface. The ratings of the two experts were averaged for each of the 14 concepts. Since each team developed two concepts—one best fit and one dark horse, the team rating for utility and novelty was obtained by considering the utility score of the best fit concept and the novelty score of the dark horse concept. We followed this approach rather than doing an average of the score of two concepts, because the design task had explicitly called for developing two concepts one higher in utility and one higher in novelty. Table 2 lists the concept scores and team scores derived from the expert ratings.

Table 2 Concept scores and team score derived from expert ratings

Team	Concepts	Concept average utility score	Concept average novelty score	Team utility score	Team novelty score
Team 1	Tree mechanism pump (dark horse)	2.95	3.2	2.6	3.2
	Pressurized u-tube (best fit)	2.6	3.55		
Team 2	Continuous sponge tube (dark horse)	2.15	3.05	3.65	3.05
	Open mine well (best fit)	3.65	1		
Team 3	Capillary pump (dark horse)	2.25	3.1	2	3.1
	Continuous sponge tube (best fit)	2	3.05		
Team 4	Continuous belt of fabric (dark horse)	2.15	3.05	2.1	3.05
	Bamboo deep lift pump (best fit)	2.1	1.9		
Team 5	Handkerchief wringer (dark horse)	2.55	2.05	3.7	2.05
	Two sequential pumps (best fit)	3.7	1		
Team 6	Handpump + Archimedes screw (dark horse)	3.1	1.35	1.35	1.35
	Solar powered propeller pump (best fit)	1.35	2.65		
Team 7	Sponge chain (dark horse)	2.45	3.4	1.75	3.4
	Opposing piston pump (best fit)	1.75	2.5		

8 Detecting Interaction Patterns Correlated with Design Outcomes

The key interaction patterns identified through analysis of the visual output were then compared with the novelty and utility ratings for the seven teams. The comparison resulted in the following findings regarding the relationship between interaction patterns and outcome measures.

1. Episodes of concept elaboration had a strong positive correlation ($r = 0.71$) with utility. The greater the number of episodes of concept elaboration in a team, the greater the utility rating of the concepts generated by the team. Figure 11 gives an example of an episode of concept elaboration identified in the visual output of the IDN Tool.

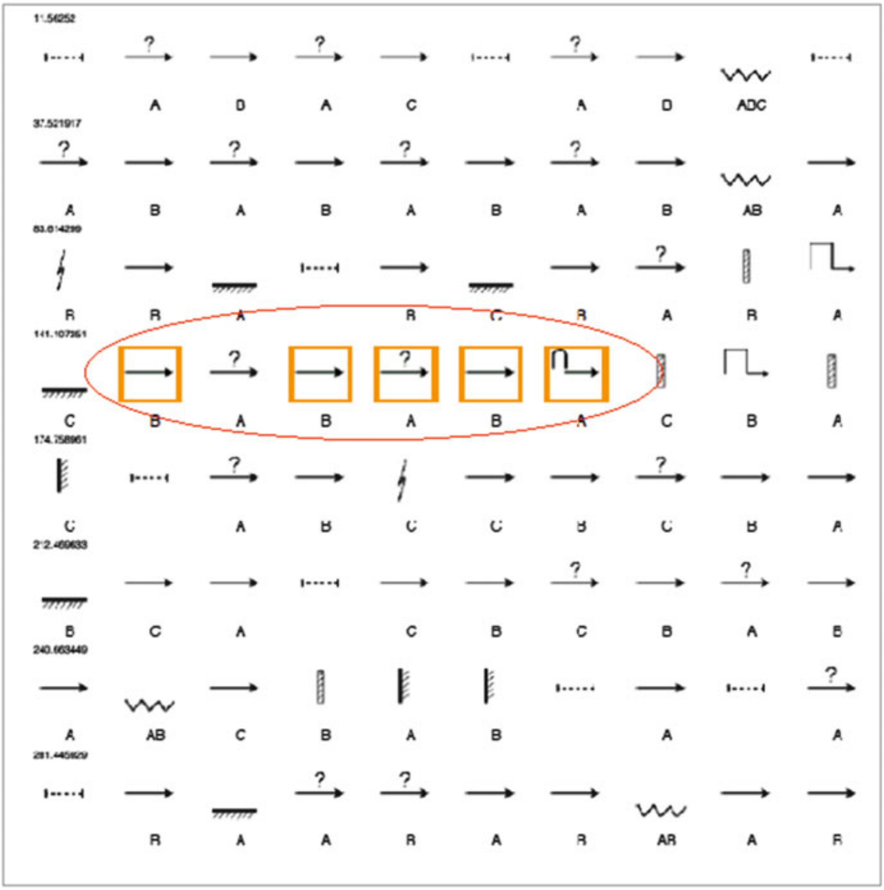


Fig. 11 This figure highlights an episode of concept elaboration. Concept elaboration episode consists of at least three consecutive idea expressions indicating that participants are contributing ideas to develop a particular solution concept

2. Dialectic episodes had a very strong positive correlation ($r = 0.9$) with utility. The greater the number of dialectic episodes in a team, the greater the utility rating of the concepts generated by the team. Figure 12 gives an example of a dialectic episode identified in the visual output of the IDN Tool.
3. Occurrence of humor had a strong positive correlation ($r = 0.55$) with utility. The greater the number of humor occurrences in a team, the greater the utility rating of the concepts generated by the team.
4. Occurrence of yes-and responses had a strong negative correlation ($r = -0.83$) with novelty. The greater the number of yes-and occurrences in a team, the lesser the novelty rating of the concepts generated by the team.

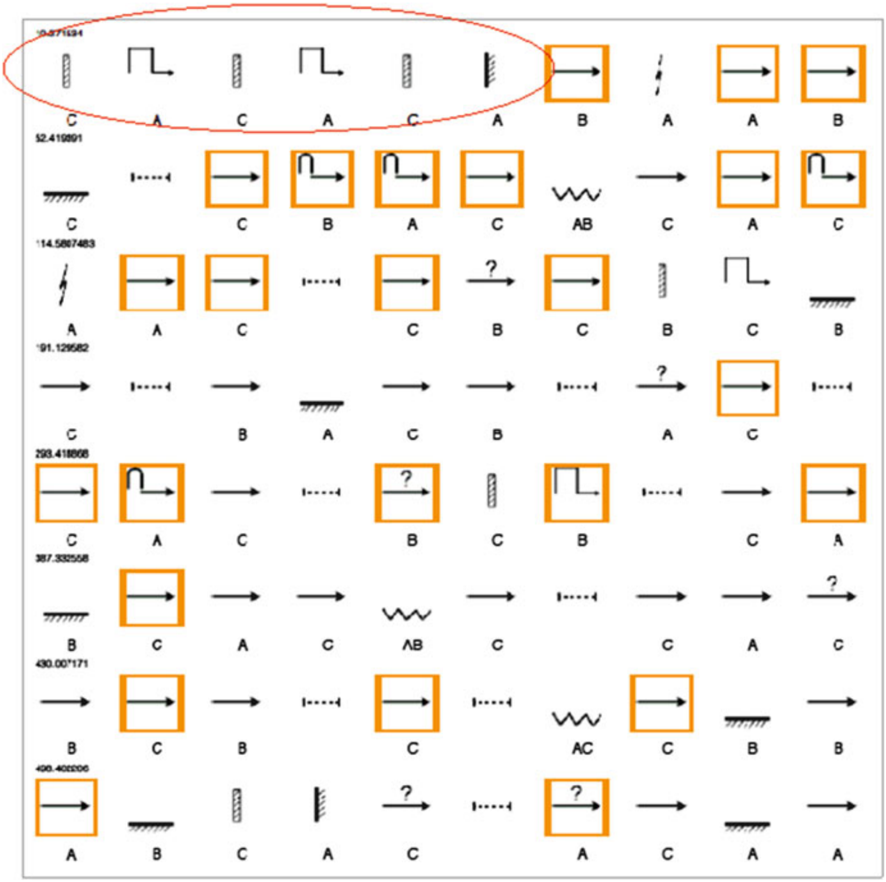


Fig. 12 This figure highlights a dialectic episode found within the visual output from the IDN Tool. A dialectic episode consists of more than two consecutive block and overcoming responses indicating that participants are engaged in an argumentative dialectic

5. Block-overcoming sequences had a moderate negative correlation ($r = -0.43$) with novelty. The greater the number of block-overcoming sequences in a team, the lesser the novelty rating of the concepts generated by the team.
6. Transitions between group work and individual work identified through occurrence of silence had a moderate positive correlation ($r = 0.34$) with novelty. The greater the number of transitions between group work and individual work in a team, the greater the novelty rating of the concepts generated by the team.

The identification of correlation between interaction patterns and design outcome measures shows how using the IDN Tool, we can detect interaction patterns that are positively or negatively associated with design outcomes. Further developments will be needed to calibrate IDN and improve its ease of use.

9 Limitations

A key limitation of the study is the small number of teams used to test the IDN Tool. Seven is a small number to obtain results that are statistically generalizable. Still, this preliminary study shows that it is possible to correlate the patterns identified through IDN Tool with design outcomes. Thus, the study can be considered significant not for its results, but for the process of establishing a protocol for testing a diagnostic instrument for design thinking teams: design activity—expert assessment of activity outcomes—pattern detection.

10 Discussion

Developing a visual diagnostic instrument for design thinking teams has implications for design thinking research, education and practice. A key element of a visual diagnostic instrument for team interaction is the reference database that indicates whether the pattern detected has any significant meaning in relation to the desired design outcome. Preparing such a database is itself a valuable research activity that could help develop a scientific foundation for our understanding of design thinking teamwork. In terms of instrumentation engineering, preparing such a database would correspond to calibrating the instrument. A key aspect of calibration with regards to design thinking is the identification and categorization of the context of the team interaction that is being analyzed. Since design is a context dependent activity, we believe that capturing the context in which team interactions occur is important in order to understand the limits of generalizability of patterns-outcome relationships. In the study of team interactions presented in this chapter, context parameters include the nature of the design activity—concept generation, the familiarity of the participants with each other—the team members were not familiar with each other before the study, and the familiarity of team members with the domain of the design brief—most participants were not familiar with the physics of water flow and pumping that formed the design challenge. We propose to repeat the study using the IDN Tool with a greater number of teams with varying context parameters in order to develop a robust calibration of relationships between interaction patterns and design outcomes mediated by the context of the design activity.

The key implication of a visual diagnostic instrument for practice and education is the capability to inform in-process feedback. Teams could be given behavioral feedback based on the visual diagnostic instrument that could enable them to improve their design performance. This feedback could be given either directly to teams or through coaches. The visual diagnostic instrument could become a coaching aid to inform coaches about on-going interaction patterns that could be conducive or detrimental to design outcomes, so that the coaches can then intervene appropriately. Thus, the development of a visual diagnostic instrument such as the IDN Tool discussed above has the potential to augment design team performance in education and practice, while being grounded in rigorous design thinking research.

11 Conclusion

In this chapter, we presented the development of a software tool called the IDN Tool that can help identify patterns of team interaction which are positively or negatively correlated with design outcomes. The IDN Tool was used to analyze concept generation interactions of seven teams. As we continue using the IDN Tool to analyze larger amounts of data, we expect to gradually build a database of team interaction patterns and their correlation with design outcomes. Such a database would form the reference database that would enable the IDN Tool to function as a diagnostic instrument for analysis of design thinking teams in industry and academia. The research presented in this chapter is the first step in that direction.

References

Carrizosa K, Eris Ö, Milne A, Mabogunje A (2002) Building the design observatory: a core instrument for design research. In: Proceedings of DESIGN 2002, the 7th international design conference, Dubrovnik

Cross N, Dorst K, Christiaans H (eds) (1996) Analysing design activity. Wiley, New York, NY

Edelman J (2011) Understanding radical breaks: media and behavior in small teams engaged in redesign scenarios. Ph.D. thesis, Stanford University

Jung M (2011) Engineering team performance and emotion: affective interaction dynamics as indicators of design team performance. Ph.D. thesis, Stanford University

Kress G (2012) The effects of team member intrinsic differences on emergent team dynamics and long-term innovative performance in engineering design teams. Ph.D. thesis, Stanford University

Pfeffer J, Sutton RI (2013) The knowing-doing gap: how smart companies turn knowledge into action. Harvard Business Press, Boston, MA

Schar M (2011) Pivot thinking and the differential sharing of information within new product development teams. Ph.D. thesis, Stanford University

Shah JJ, Smith SM, Vargas-Hernandez N (2003) Metrics for measuring ideation effectiveness. Des Stud 24(2):111–134

Sonalkar N (2012) A visual representation to characterize moment-to-moment concept generation through interpersonal interactions in engineering design teams. Ph.D. thesis, Stanford University

Sonalkar N, Mabogunje A, Leifer L (2013) Developing a visual representation to characterize moment-to-moment concept generation in design teams. Int J Des Creat Innov 1(2):93–108

Tang J, Leifer L (1991) An observational methodology for studying group design activity. Res Eng Des 2(4):209–219

Törlind P, Sonalkar N, Bergström M, Blanco E, Hicks B, McAlpine H (2009) Lessons learned and future challenges for design observatory research. In: Proceedings of ICED 09, the 17th international conference on engineering design, vol 2, Design theory and research methodology, Palo Alto, CA, USA

Valkenburg R (2000) The reflective practice in product design teams. Delft University of Technology, Delft

Design Thinking Health: Telepresence for Remote Teams with Mobile Augmented Reality

Lauren Aquino Shluzas, Gabriel Aldaz, and Larry Leifer

Abstract This research examines the capabilities and boundaries of a hands-free mobile augmented reality (AR) system for distributed healthcare. We use a developer version of the Google Glass™ head-mounted display to develop software applications to enable remote connectivity in the healthcare field; characterize system usage, data integration, and data visualization capabilities; and conduct a series of pilot studies involving medical scenarios. This book chapter discusses the need for a AR head-mounted display to improve chronic wound care photography and to facilitate surgical interventions. We provide an overview of the system architecture used in this research, and highlight future applications of AR systems for improved clinical care.

In the broader context of distributed collaboration for improved healthcare delivery, this research provides a foundation for: (i) examining the use of technology for complex distributed problem solving through interdisciplinary collaboration; (ii) gaining an improved understanding of the benefits of human augmentation through enhanced visualization and auditory capabilities, on healthcare team performance; and (iii) exploring an AR system's ability to influence behavior change in situations requiring acute decision-making through interaction between centralized experts and point-of-impact delivery personnel. Moreover, this chapter provides insight into the need for future IT systems engineering projects aimed at enhancing healthcare connectivity and distributed care.

Abbreviations

AR	Augmented reality
CMS	Centers for medicare and medicaid services
DSLR	Digital single-lens reflex
EMR	Electronic medical record
HMD	Head-mounted display
HAPUs	Hospital-acquired pressure ulcers
ICU	Intensive care unit
IMU	Internal measurement unit

L.A. Shluzas (✉) • G. Aldaz • L. Leifer
Center for Design Research, Building 560, 424 Panama Mall, Stanford, CA 94305-2232, USA
e-mail: lauren.aquino@stanford.edu; zamfir@stanford.edu; leifer@cdr.stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_5

MRN	Medical record number
MOWA	Mobile wound analyzer
NPUAP	National pressure ulcer advisory panel
P2P	Peer-to-peer
VN	Virtual nurse

1 Introduction

In the healthcare field, the need for improved tools to enhance collaboration among patients and providers has become increasingly urgent—due, in large part, to a global rise in aging populations and chronic disease prevalence, coupled with increasing health care costs and physician shortages worldwide (Mattke et al. 2010; Magnusson et al. 2004). To address these challenges, this research examines the use of wearable mobile computing to mediate interdisciplinary communication and collaboration in healthcare. We present an overview of collaborative technologies and interaction modalities for home-healthcare and hospital use. Advances in mobile computing include technologies to improve remote patient monitoring, automate simple knowledge-base procedures, and facilitate the delivery of interventions.

1.1 Remote Patient Monitoring

For home healthcare applications, existing systems enable medical staff to remotely monitor patients suffering from advanced chronic disease and provide prompt support regarding health education and treatment compliance. The *Vsee* video collaboration system (Sunnyvale, CA), for instance, simplifies patient-doctor interactions through web-based video calling, coupled with medical device integration.

In the consumer health and wellbeing space, a variety of products provide self-patient monitoring to encourage behavior modifications aimed at promoting healthier lifestyles. These include activity monitors such as the *Nike Fuelband* and *Fitbit Flex*, as well as the *LUMOback* real-time posture feedback system. Yet, despite current advances, home monitoring technologies are often limited by the dependency on appropriate bandwidth, customized networks and high-cost equipment, as well as a lack of integration into electronic medical record (EMR) systems.

In the hospital setting, remote monitoring systems, such as the *tele-ICU*, aid clinicians in the delivery of care to ICU patients (Fig. 1). By collaborating with the bedside team, the *tele-ICU* silently assists in the delivery of timely interventions (Goran 2010). Yet, financial barriers associated with installing and operating such systems has limited widespread adoption.

Fig. 1 Tele-ICU: High-resolution cameras mounted in each ICU unit (Goran 2010)

1.2 Automating Knowledge-Based Procedures

For patient-virtual agent interaction, the animated virtual nurse (VN) is an automated system that teaches patients their post-discharge self-care regimen directly from their hospital beds (Bickmore et al. 2009). This system incorporates a VN who embodies best practices in health communication for patients with inadequate health literacy, and illustrates a growing field in mobile computing aimed at increasing universal healthcare access (Fig. 2).

Cognitive aids, such as dynamic checklists, present another example of tools to facilitate collaboration in the clinical setting, through automating knowledge-based procedures. A recent study involving dynamic checklists found that medical crisis care situations reveal “a physically complex information space, and relatively high-tempo time scales of gaze, action, and team-based coordination and communication” (Wu et al. 2011).

1.3 Facilitating the Delivery of an Intervention

Collaborative technologies that facilitate the delivery of an intervention may include those in which (i) a clinician (expert) aids a non-expert in delivering an intervention, and (ii) a clinician delivers an intervention remotely through a robotic interface. With the commercialization of high-speed data networks, the implementation of these scenarios may be realized through the use of augmented reality (AR) systems. Many AR applications provide the benefit of visualizing three-dimensional data captured from non-invasive sensors, and range from remote 3D image analysis to advanced telesurgery (van Krevelan and Poelman 2010; Zhou et al. 2008).

Despite existing technologies, however, there is a growing need for new tools capable of augmenting a clinician’s knowledge base and his/her complex problem

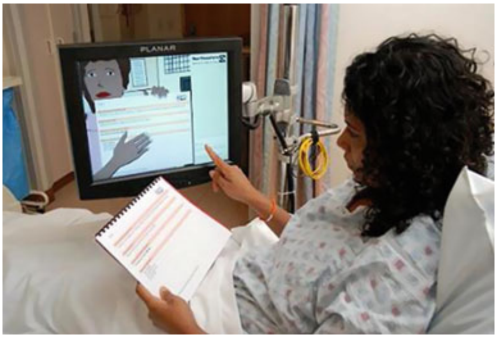


Fig. 2 Virtual nurse agent that conducts a bedside dialogue with patients immediately prior to hospital discharge (Bickmore et al. 2009)

solving ability, while performing an intervention. Such augmentation can be accomplished if a system simultaneously connects the clinician (expert) to relevant medical databases, other experts, and a live telemetry of patients' vital statistics. The first-order challenge required to accomplish these connections is the ability to manage the resulting high-bandwidth information flow between human and computer agents, and to enable agents to collectively work as a design team.

To address these challenges, this ongoing research effort examines the use of hands-free mobile AR for distributed healthcare collaboration. We hypothesize that a mobile AR system (head-mounted display) will shorten communication cycles and reduce errors associated with point-of-care decision-making and distributed collaboration in healthcare scenarios.

To test our hypothesis, we obtained four pairs of Google Glass™ as a platform for research. Our methodology involves (i) clinical needs finding to ground the study in the context of high-impact clinical problems; (ii) software development to create customized applications for a head-mounted display that are specific to two or more clinical areas; and (iii) pilot testing to characterize the AR system's usage, data integration, and data visualization capabilities. In this chapter, we focus on the need for a head-mounted display to improve chronic wound care photography and to facilitate surgical interventions. We provide an overview of the system architecture used in this research, and highlight future applications of AR systems for improved clinical care.

2 Clinical Needs Finding

To narrow our research focus, we began with needs finding at Stanford Health Care from October to December 2013. We interviewed ten nurses and two physicians, and shadowed four additional nurses ($n = 16$ clinicians), to generate over 135 needs. We grouped needs into 15 broad clinical areas and ranked each category on a 5-point scale based on degree of importance to the hospital (pain point), alignment with research interests, and feasibility (Aldaz et al. 2015) (Table 1).

Table 1 Clinical needs finding, top needs based on a 5-point scale (Aldaz et al. 2015)

Clinical area	Target application areas	Degree of importance to the hospital (Pain point)	Alignment with research interests	Feasibility	Mean score
Documentation (Video recordings/ Photographs)	Wound care assessment	4	4	5	**4.3**
Documentation (EMR Data)	Wound care assessment	5	3	4	**4.0**
Alerts and alert fatigue	Surgical monitoring and Response to cardiac arrest	5	3	4	**4.0**
Improving live visualization of cases	Point-of-view sharing during surgery	3	3	5	**3.7**
Communication	Multiple areas	5	2	4	**3.7**

Within the top-ranking categories, we segmented needs by degree of clinical risk to patients. Our top three needs (in order of low to high patient risk) included: wound and skin care photography, point-of-view sharing during surgery, and vital sign communication during cardiac arrest (Shluzas et al. 2014a, b). We selected chronic wound photography as an initial target focus area since: (i) a reduction in the incidence of chronic wounds—especially hospital-acquired pressure ulcers—is of paramount concern to healthcare facilities; and (ii) chronic wound image capture involves a relatively low degree of clinical risk for patients, thus enabling a solution to be tested and implemented quickly (Aldaz et al. 2015).

3 The Need for an AR System in Wound Care

3.1 A Head-Mounted Display for Hospital-Acquired Pressure Ulcers

Chronic wounds affect 6.5 million Americans and pose a $25 billion annual financial burden to the U.S. health care system (Sen et al. 2009). Of particular importance to hospitals are pressure ulcers (bedsores), which along with venous and diabetic ulcers, comprise the vast majority of chronic wounds. A pressure ulcer is "localized injury to the skin and/or underlying tissue usually over a bony prominence, as a result of pressure, or pressure in combination with shear. A number of contributing or confounding factors are also associated with pressure ulcers; the significance of these factors is yet to be elucidated" (NPUAP 2007). Studies have reported that a pressure ulcer extends the length of stay for an acute hospital admission by 7–50 days (Graves et al. 2005). Nearly 60,000 hospital patients in the U.S. die each year from complications due to hospital-acquired pressure ulcers (HAPUs), while the direct treatment costs for HAPUs at an average hospital range between $400,000 and $700,000 (Gonzales and Pickett 2011). In 2007, the Centers for Medicare and Medicaid Services (CMS) made a policy decision not to pay for the most severe HAPUs (Lyder et al. 2012). Thus, to reduce the incidence of HAPUs, hospitals are mandating thorough and accurate documentation, which is necessary to determine the degree of wound severity, evaluate the effectiveness of therapies, and modify treatment plans as appropriate.

Increasingly, healthcare facilities are supplementing direct observation with digital photography for assessment and documentation (Ahn and Salcido 2008). The National Pressure Ulcer Advisory Panel (NPUAP) strongly encourages that every wound photograph include patient identification, date and time markings, and wound dimensions (NPUAC 2014). In practice, nurses write each patient's personal identification (ID) information—e.g. name and medical record number (MRN)—on a paper ruler that they hold next to the wound and include in the picture frame (Fig. 3).

Nurses commonly use digital cameras for wound photography, but the process is often tedious, time consuming, and not private. At Stanford Healthcare, uploading wound images to a patient's Electronic Medical Record (EMR) typically requires up to 2 days. Recent advances in mobile technology have significantly reduced the time between taking wound photographs and uploading images to a patient's EMR. Nonetheless, smartphones and tablets still have a significant drawback for wound photography: they are hands-on devices. Consequently, two to three nurses are required to capture wound images—one to hold a camera, a second to hold a ruler and position the patient, and often a third nurse to assist with patient positioning. Furthermore, the physical handling of cameras increases the likelihood of cross-contamination and patient infections (Aldaz et al. 2015).

Fig. 3 Nurse writing patient information on a disposable paper ruler

3.2 Existing Mobile Devices for Wound Photography

The system under development builds on prior work in mobile devices for wound photography. The types of devices capable of non-contact wound photography include general-purpose devices (e.g., digital cameras, smartphones, and tablets), wearable (e.g., optical head-mounted displays), and special-purpose devices.

General-Purpose Devices A common way for nurses to take wound photographs is via dedicated digital cameras (Hayes and Dodds 2003). After taking photos, a nurse transfers images to a secure, password-protected computer, and deletes the images from the camera. Since digital cameras are not networked, a nurse would then transfer the images via a physical memory card. At Stanford Healthcare, nurses insert a memory card in a printer to make physical prints. He or she then places each print on a separate piece of paper and fills in relevant information with a pen. Each day, couriers transport a stack of these documents offsite for scanning. It requires up to 2 days for the images to be uploaded to a patent's EMR.

Epic Systems (Verona, WI) is a leading provider of EMR software (Epic Systems 2014). The Epic Haiku mobile application allows nurses to take photographs using a general-purpose smartphone and immediately upload images to a patient's EMR, where other clinicians may view the images (Beard and Hamid 2014). Photos taken within the Epic Haiku app are not stored on the phone's memory, thereby ensuring compliance with security and privacy protocols (Beard and Hamid 2014). To use Epic Haiku, a nurse typically holds the smartphone in both hands while another nurse holds the paper ruler and positions the patient. Alternatively, one nurse can hold the paper ruler in one hand and the smartphone in the other, in which case pressing the button to take a photograph is often challenging (Fig. 4). After taking each image, the nurse must manually enter a text description, which requires one's full attention and creates a barrier between the

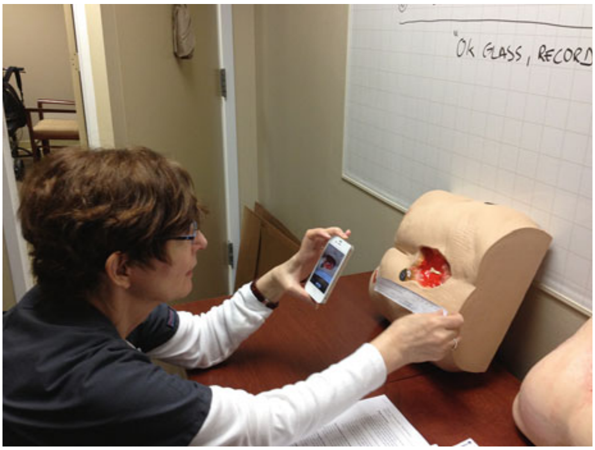


Fig. 4 Nurse taking a digital image using the iphone Epik Haiku application and a paper ruler

nurse and patient. After the patient encounter, the nurse meticulously wipes down all smartphone surfaces to disinfect the device.

MOWA (Mobile Wound Analyzer) is an application for Android smartphones and tablets that takes photos and uses software-based technologies to calculate wound area and identify three types of tissues in the bed of the lesion (Mobile Wound Analyzer 2014). Presently, however, MOWA does not integrate with EMRs, thus limiting its utility.

Wearable and Special-Purpose Devices In addition to hand-held digital cameras, a range of studies has demonstrated the use of head-mounted cameras (e.g. GoPro Hero3) for clinical use and education (Bizzotto et al. 2013). A major advantage of wearable head-mounted cameras, "is that the user is not restricted in his or her manual and oculomotor actions by camera operation tasks" (Wagner et al. 2006). This allows the performance of actions in a natural way with increased mobility. Although often WiFi-enabled for remote viewing, wearable cameras typically lack patient identification and EMR integration capabilities for routine clinical care.

Special-purpose devices usually include sophisticated wound-measurement methods, in addition to photographic capabilities. These devices achieve increased accuracy through vision-based technologies such as stereophotography or structured lighting. For example, MAVIS III (Perry Baromedical, Riviera Beach, FL) is a stereoscopic, hand-held camera that measures the area, volume, and circumference of chronic wounds (MAVIS III 2014). Traditionally, special-purpose devices have suffered from one or more of the following drawbacks: they may be expensive, cumbersome to use in a clinical setting, or require significant training time (Krouskop et al. 2002).

4 The Need for Improved AR Systems in Surgery

Beyond mobile devices for wound care photography, augmented reality "mediates ideas between humans and computers, humans and humans, and computers and humans" (Craig 2013). Such systems can supplement the real world with virtual objects that appear to coexist in the same space (Zhou et al. 2008), and have the potential to provide contextual, situated learning experiences for users (Zhu et al. 2014). With a wearable head-mounted display, the computer can unobtrusively display messages to gain a user's attention, as well as display text or graphics in one's visual field (Starner et al. 1997).

Wearable augmented reality (AR) systems have been demonstrated in a range of clinical application areas, such as image-guided surgery, pre-operative imaging training, health-related behavior change, and clinical education (Craig 2013; Dixon et al. 2013; Doherty et al. 2013). In particular, a range of studies in the clinical field has demonstrated the use of wearable technology with cameras and optical head-mounted displays. Bizzotto et al. (2013), for instance, used the GoPro HERO3 in percutaneous, minimally invasive, and open surgeries with high image quality and resolution. Likewise, though not widely distributed for consumer use (as of January 2015), a number of proof-of-concept projects and studies using the Google Glass™ optical head-mounted display (Google Glass 2013) have been demonstrated in the medical field. The Glass technology can reflect projected images while allowing the user to see through the device's eyepiece. Medical applications for Glass have included remote collaboration between doctors, heads-up viewing of vital signs, medications, and lab reports, as well as live streaming of operations to medical students (Glauser 2013). Several reports document the live two-way broadcasts of actual patient surgeries by doctors wearing Glass in the operating room (Lutz and Kwan 2014; Medical Xpress 2013).

Although studies have examined the use of wearable AR systems for both in-patient and remote medical procedures (Lutz and Kwan 2014; Medical Xpress 2013; Dixon et al. 2013), there is a growing need for the development of integrated systems that enable users to seamlessly capture and annotate visual images in a hands-free manner, and integrate this information with a patient's EMR. Outside of wound care, a specific application in which hands free image capture is particularly critical includes point-of-view (POV) data sharing during surgery.

The development of a software application for POV sharing during surgery requires three key elements: voice commands, bi-directional communication, and EMR data transfer. The use of voice commands to control digital image capture in a hands-free manner is particularly useful for clinicians to maintain heightened sterility while performing surgical procedures. Bi-directional communication is required to enable collaboration between surgeons wearing a head-mounted display and remote colleagues, as well as to establish connectivity with remote sensors in the healthcare environment.

In addition to software development, the physical augmentation of head-mounted displays is needed to bring wearable technology and wearable computing to the operating room environment. Physical enhancements include, for instance, a transparent splash shield that surgeons may adhere to the front frame of a head-mounted display for protection from infectious disease, an optical loupe mounted to the frame (in front of a surgeon's eye) in order to increase magnification for surgical procedures, and elements encased on an HMD in a protective cover for improved cleaning and robustness during routine clinical use.

5 Current System Development

To address the needs for improved hands-free visualization, image capture, and EMR transfer, we are currently developing a Google Glass application (known as "Glassware") and an Android smartphone application. This application was initially designed to be used by nurses for chronic wound photography (Shluzas et al. 2014a, b; Aldaz et al. 2015), but is now extending into surgery and other clinical areas. Google Glass was selected as an initial platform for research, since it has an optical head-mounted display that is capable of taking pictures and recording videos using an integrated camera. The device also has the ability to communicate wirelessly via WiFi and Bluetooth. Additional sensors integrated into the device enable the development of augmented reality-based applications. We developed Glassware applications using the Glass Development Kit (GDK) for Android 4.0.4 (API Level 15, Ice Cream Sandwich), and the smartphone app for Android 4.3 (API Level 18, Jelly Bean) (Shluzas et al. 2014a, b).

The novel features of the system include (i) barcode scanning using the Glass camera, and tagging subsequent images with a patient's personal identification information that is embedded in the barcode, (ii) capturing a live video preview in the Glass eyepiece before a photo is taken, (iii) using a double-blinking gesture to take photographs, utilizing Glass' IR sensor, and (iv) using a head-tilt gesture (while in the preview mode) to zoom in and out of an image, and a head-tilt gesture to send images to a patient's EMR, through the use of Glass' internal measurement unit (IMU) sensor (Aldaz et al. 2015).

5.1 *Google Glass Sensors*

Google Glass has four sensors that we have used to implement the system for hands-free wound photography: a camera, an inertial measurement unit (IMU), an infrared sensor facing toward the user's right eye, and a microphone (Fig. 5).

Camera with Digital Zoom Albrecht et al. (2014) concluded that Glass was efficient for acquiring images, but for photo documentation in forensic medicine, the image quality was inferior to those from a Digital Single-Lens Reflex (DSLR) camera. This finding is perhaps not surprising, considering that Google Glass uses a standard 5-megapixel (2528 × 1956) smartphone camera with a fixed focal length of 3 mm. The Glass camera has a wide-angle lens that captures a field of view too

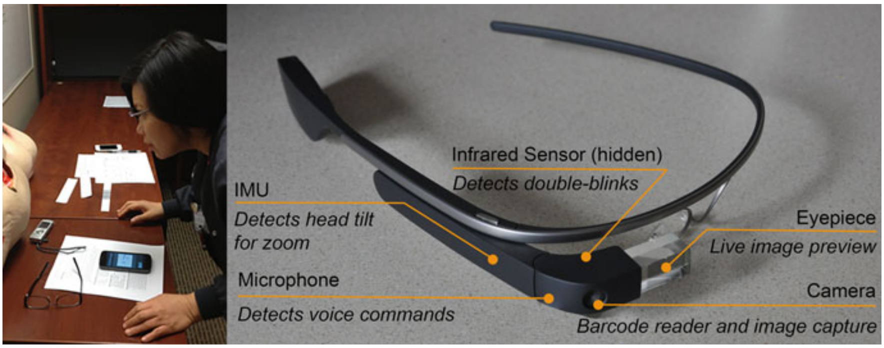


Fig. 5 System leveraging the camera and internal sensors of the Glass™ wearable computing device to create a mobile hands-free application for capturing, tagging, and transferring digital images to a patient's medical record (Aldaz et al. 2015)

large for most typical medical applications and illustrates the need for zoom (Muensterer et al. 2014).

Because Glass has a fixed focus, all zooming is digital and does not lead to a gain in optical resolution. In theory, a photo could be cropped in post-processing to obtain an identical image as one taken after digital zooming, but our preliminary research showed that nurses appreciated having the ability to zoom before taking the picture. In practice, we discovered one drawback with zooming before picture taking. Usually, Glass uses burst-mode photography (capturing a rapid sequence of shots to improve dynamic range or reduce noise) when instructed to take a picture (Google + Glass Update 2014). Burst-mode is lost when taking a picture after a zoom, resulting in lower quality photos.

Barcode Reader Hospital patients wear ID wristbands printed with a medical record number and corresponding barcode. We implemented a barcode reader based on the open source, cross-platform library ZBar (2014). ZBar works very well with newer smartphones, like the iPhone 5, that have auto-focusing cameras. Since the Google Glass built-in camera has a fixed focus, we used trial and error to find a barcode size that ZBar could read at arm's length, where the image was focused (we found that objects closer than approximately 30 cm would not be in focus). Furthermore, we had to use a flat barcode, rather than the curved barcode on wristbands used at most hospitals.

Microphone The Android 4.0.4 GDK version allows developers to designate a custom voice command to launch an application. The Glassware digital image capture applications we developed is launched with the command, "*ok Glass, take clinical image*." We also used the microphone for wound annotation purposes, whereby our app sent audio to the Google servers for speech-to-text conversion. Previous investigators were surprised to find that Glass recognized complex medical terms such as "Microvillous Inclusion Disease" about half the time, but cautioned that for reliable medical use, however, the error rate had to decrease substantially (Muensterer et al. 2014).

IMU and Infrared Sensor Ishimaru et al. (2014), who argued that the inertial measurement unit (IMU) and blink detection are the most characteristic features of the Google Glass platform, studied combining head motion with blink frequency for activity recognition. For the our research and testing purposes, the IMU and infrared sensor each provided a separate input for hands-free interaction. We used the IMU (gyroscope and accelerometer) to measure head tilt and the infrared sensor to detect double blinks.

6 Conclusion

This research examines the use of augmented reality technologies to enhance complex problem solving and distributed care for a range of clinical application areas. Specifically, we provide a rationale for the development of a hands-free image capture system in the wound care field. Our initial pilot study revealed that the system enables hands-free digital photography, tagging, speech-to-text image annotation, and the transfer of data to an electronic medical record. In its current implementation, the system leverages Google Glass' camera and internal sensors to guide clinicians through the process of taking and annotating wound images.

Additional applications include augmented point-of-view sharing during operating room procedures to enhance a clinician's knowledge base and decision-making capabilities, while performing surgical interventions. Specifically, since several individuals are needed to perform a surgical procedure (e.g. the attending surgeon, assisting surgeons, an anesthesiologist, and multiple nurses), the incision site is frequently crowded and obstructed from one's field-of-view. With augmented POV sharing, images projected from an attending surgeon's vantage point onto a remote screen provide visual clarity to individuals directly involved in the surgery, as well as to remote participants and expert advisors. It also helps trainees and surgical fellows to learn procedures more accurately, by viewing an intervention from the same perspective as an attending surgeon, rather than from the reverse perspective (e.g. standing on the opposite side of the table)—which is often the case today. Finally, the use of POV sharing with bi-directional communication capabilities and remote sensor connectivity enables real-time collaboration with a pathology lab or expert consultants while performing a surgical intervention. Through visual overlays, POV sharing can co-locate one's visual field with information critical to performing a procedure—such as vital sign information, procedural descriptions, or MRI scans.

In the future, we aim to further enhance the robustness and reliability of a mobile AR system for acute care scenarios in the operating room and emergency room settings. This involves developing applications with data security and privacy features that are in compliance with strict hospital security protocols. We are examining the use of a dynamic digital ruler to measure images in a hands-free manner, a method of contextual time-lapse image recall using the Glass display to facilitate nurses in visualizing a wound's staging and healing progression over time, and speech-to-text voice recognition capabilities for improved wound annotation and team-based collaboration.

In the broader context of distributed collaboration for improved healthcare delivery, this research aims to examine the use of technology for complex distributed problem solving through interdisciplinary collaboration; gain an improved understanding of the benefits of human augmentation through enhanced visualization and auditory capabilities, on healthcare team performance; and explore an AR system's ability to influence behavior change in situations requiring acute decision-making through interaction between centralized experts and point-of-impact delivery personnel.

References

Ahn C, Salcido RS (2008) Advances in wound photography and assessment methods. Adv Skin Wound Care 21:85–93
Albrecht UV, von Jan U, Kübler J, Zoeller C et al (2014) Google glass for documentation of medical findings: evaluation in forensic medicine. J Med Internet Res 16(2), e53
Aldaz G, Shluzas LA, Pickham D, Eris O, Sadler J, Joshi S, et al. (2015) Hands-Free Image Capture, Data Tagging and Transfer Using Google Glass: A Pilot Study for Improved Wound Care Management. PLoS ONE 10(4):e0121179. doi: 10.1371/journal.pone.0121179
Beard HR, Hamid KS (2014) Worth a thousand words: integrating clinical photographs into an electronic medical record. Healthcare 2(1):22–25
Bickmore TW, Pfiefer LM, Jack BW (2009) Taking time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the CHI 2009. ACM Press
Bizzotto N, Sandri A, Lavini F, Dall'oca C, Regis D (2013) Video in operating room: GoPro HERO3 camera on surgeon's head to film operations—a test. Surg Innov. doi: 10.1177/1553350613513514
Craig AB (2013) Understanding augmented reality: concepts and applications. Elsevier, Boston, MA
Dixon BJ, Daly MJ, Chan H, Vescan AD, Witterick IJ, Irish JC (2013) Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. Surg Endosc 27:454–461
Doherty A, Williamson W, Hillsdon M, Hodges S, Foster C, Kelly P (2013) Influencing health-related behaviour with wearable cameras: strategies & ethical considerations. In: Proceedings of the 4th International SenseCam & Pervasive imaging conference. ACM, New York, NY, pp 60–67
Epic Systems. http://www.epic.com. Accessed 10 Jan 2014
Glauser W (2013) Doctors among early adopters of Google Glass. Can Med Assoc J 185:1385
Gonzales D, Pickett R (2011) Too much pressure. Stanford Nurse 31(1):12–13
Google + Glass Update. http://plus.google.com/+GoogleGlass/posts/9CdE6ysitvP. Accessed 21 Jan 2014
Google Glass. http://www.google.com/glass/start. Accessed 5 Aug 2013
Goran SF (2010) A second set of eyes: an introduction to Tele-ICU. Crit Care Nurse 30(4):46–55
Graves N, Birrell F, Whitby M (2005) Effect of pressure ulcers on length of hospital stay. Infect Control Hosp Epidemiol 26:293–297
Hayes S, Dodds S (2003) Digital photography in wound care. Nurs Times 99(42):1–3
Ishimaru S, Weppner J, Kunze K, Kise K, Dengel A, Lukowicz P, Bulling A (2014) In the blink of an eye: combining head motion and eye blink frequency for activity recognition with Google Glass. In: Proceedings of the 5th augmented human international conference. pp 150–153
Krouskop TA, Baker R, Wilson MS (2002) A noncontact wound measurement system. J Rehabil Res Dev 39:337–345
Lutz BJ, Kwan N (2014) Chicago surgeon to use Google Glass in operating room. http://www.nbcchicago.com/news/tech/google-glass-surgery-237305531.html. Accessed 2 Feb 2014
Lyder CH, Wang Y, Metersky ML et al (2012) Hospital-acquired pressure ulcers: results from the national medicare patient safety monitoring system study. J Am Geriatr Soc 60:1603–1608
Magnusson L, Hanson E, Borg M (2004) A literature review study of information and communication technology as a support for frail older people living at home and their family carers. Technol Disabil 16:223–235
Mattke S, Klautzer L, Mengistu T, Garnett J, Hu J, Wu H (2010) Health and well-being in the home: a global analysis of needs, expectations, and priorities for home health care technology. RAND Corporation, Santa Monica, CA
MAVIS III. http://www.perrybaromedical.com/accessories-mavis3.html. Accessed 7 Apr 2014
Medical Xpress. First US surgery transmitted live via Google Glass (with video http://medicalxpress.com/news/2013-08-surgery-transmitted-google-glass-video.html. Accessed 13 Oct 2013

Mobile Wound Analyzer. http://www.healthpath.it/mowa.html. Accessed 6 Apr 2014

Muensterer OJ, Lacher M, Zoeller C, Bronstein M, Kübler J (2014) Google Glass in pediatric surgery: an exploratory study. Int J Surg 12:281–289

National Pressure Ulcer Advisory Council (2014) FAQ: photography for pressure ulcer documentation. http://www.mnhospitals.org/Portals/0/Documents/ptsafety/skin/NPUAP-photographyFAQ.doc. Accessed 16 Mar 2014

NPUAP Pressure Ulcer Stages/Categories (2007) http://www.npuap.org/resources/educational-and-clinical-resources/npuap-pressure-ulcer-stagescategories/. Accessed 11 Nov 2014

Sen CK et al (2009) Human skin wounds: a major and snowballing threat to public health and the economy. Wound Repair Regen 17:763–771

Shluzas LA, Aldaz G, Sadler J, Joshi SS, Leifer L (2014a) Mobile augmented reality for distributed healthcare: point-of-view sharing during surgery. In: Proceedings of the fourth international conference on ambient computing, applications, services and technology (Ambient 2014), IARIA, pp 34–38

Shluzas LA et al (2014b) A wearable computer with a head mounted display for hands-free image capture and bi-directional communication with external sensors. US Provisional Patent No. 61968857, Mar 2014

Starner T, Mann S, Rhodes B, Levine J, Healey J, Kirsch D, Picard R, Pentland A (1997) Augmented reality through wearable computing. Presence 6(4):386–398

van Krevelan DWF, Poelman R (2010) A survey of augmented reality technologies, applications, and limitations. Int J Virtual Real 9(2):1–20

Wagner P, Bartl K, Gunthner W, Schneider E, Brandt T, Ulbrich H (2006) A pivotal head mounted camera that is aligned by three dimensional eye movements. In: Proceedings of the 2006 symposium on eye tracking research & applications. ACM, New York, NY, pp 117–124

Wu L, Cirimele J, Card S, Klemmer S, Chu L, Harrison K (2011) Interactive cognitive aids: maintaining shared mental models in anesthesia crisis care with nurse tablet input and large-screen displays. UIST'11, ACM

ZBar barcode reader. http://zbar.sourceforge.net/. Accessed 12 Feb 2014

Zhou F, Duh HBL, Billinghurst M (2008) Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR. In: Proceedings of the 7th IEEE/ACM international symposium on mixed and augmented reality. ISMAR, pp 193–202

Zhu E, Hadadgar A, Masiello I, Zary N (2014) Augmented reality in healthcare education: an integrative review. doi:10.7287/peerj.preprints.335v1

Talkabout: Making Distance Matter with Small Groups in Massive Classes

Chinmay Kulkarni, Julia Cambre, Yasmine Kotturi, Michael S. Bernstein, and Scott Klemmer

Abstract Massive online classes are global and diverse. How can we harness this diversity to improve engagement and learning? Currently, though enrollments are high, students' interactions with each other are minimal: most are alone together. This isolation is particularly disappointing given that a global community is a major draw of online classes. This paper illustrates the potential of leveraging geographic diversity in massive online classes. We connect students from around the world through small-group video discussions. Our peer discussion system, Talkabout, has connected over 5000 students in 14 online classes. Three studies with 2670 students from two classes found that globally diverse discussions boost student performance and engagement: the more geographically diverse the discussion group, the better the students performed on later quizzes. Through this work, we challenge the view that online classes are useful only when in-person classes are unavailable. Instead, we demonstrate how diverse online classrooms can create benefits that are largely unavailable in a traditional classroom.

1 Introduction

At their best, culturally diverse classrooms leverage students' different backgrounds to improve learning and foster cultural understanding. When students engage with peers from different cultures, they become aware of their own assumptions and how others have different perspectives (Nemeth 1986). This shifts students from 'automatic' thinking to more 'active, effortful, conscious' thinking,

C. Kulkarni (✉) • M.S. Bernstein
Stanford University, Stanford, CA, USA
e-mail: chinmay@cs.stanford.edu; msb@cs.stanford.edu

J. Cambre
Stanford University and Coursera Inc., Stanford, CA, USA
e-mail: jcambre@cs.stanford.edu

Y. Kotturi • S. Klemmer
UC San Diego, La Jolla, CA, USA
e-mail: ykotturi@ucsd.edu; srk@ucsd.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_6

which aids learning and growth (Gurin et al. 2002). But, while physical classrooms often strive to be diverse, they remain limited by physical geography (Kucsera and Orfield 2014).

Massive online courses recruit thousands of students from over 100 countries, bringing together peers with many nationalities and experiences (Olds 2013). Instructors often advertise how many countries are represented in the class (Breslow et al. 2013; Konstan et al. 2014; Olds 2013). However, while student diversity has become a calling card of online education, this potential is currently un-tapped. Most online students currently see only a glimpse of their peers' global diversity, primarily in text discussion forums. This slow-motion communication is a poor fit for the open-ended dialogue characteristic of dorm hallway conversation (Jacobs 2013), and can reinforce a one-size-fits-all, broadcast educational approach (Losh 2014).

This paper illustrates the potential of leveraging diversity in online classes, and introduces the Talkabout environment and curricula for small, geographically-diverse groups in massive classes. Talkabout connects students to their global peers via guided, synchronous video discussion. Talkabout focuses on harnessing geographic diversity, where students connect with peers from other parts of the world. Geographic diversity enables students to access peers with different cultures (Green 2005), levels of income (Goesling 2001), and beliefs about learning (Purdie and Hattie 1996).

Geographically diverse classrooms can improve educational experiences, making them deeper and more realistic. Multinational discussions create the opportunity for what one student called a 'mini United Nations', where students experience first-hand the differing concerns and beliefs of people from different countries.

Talkabout forms groups of two to nine students from different parts of the world for a video discussion. Discussion prompts ask peers to relate course content to their local and personal experiences, encouraging students to reflect on previously unexamined assumptions about their own environments, and deepening their learning (Lin and Schwartz 2003). To date, more than 5000 students from 134 countries have used Talk-about in 14 online classes via Coursera and OpenEdX. This paper reports results from the first seven courses and 3200 students. These classes included Social Psychology, Organizational Analysis, Behavioral Economics, and Logic and Design. Table 1 shows a sampling of topics discussed. The median discussion had six students from five countries.

Talkabout's discussion sessions improved student engagement: students randomly assigned to a Talkabout group were significantly more likely to participate in class quizzes than those placed on a wait-list for future participation (Wald $z^* = 1.96$, $p = 0.03$).

Geographically diverse discussions yield higher grades and engagement. A controlled experiment in two massive online classes varied the number of countries present in Talkabout discussions. Students in more geographically diverse discussions performed significantly better on subsequent quizzes and exams ($t(129) = 1.78$ and $t(110) = 2.03$, $p < 0.05$).

Table 1 Excerpts from discussion agendas from 1 week in different classes. Each question below included more detailed guidance in the actual discussion

Course title	Representative discussion topics
Critical Perspectives on Management	How do you define innovation and invention? How do manage them? Are shipping containers and labor unions innovations or inventions?
Irrational Behavior	How do you treat money as a relative rather than absolute good? Do you think that it is more painful to pay with cash than credit? How might issues of fairness vary by culture?
Organizational Analysis	Describe your experience in organizations where decisions by organized anarchy occurred. Did they solve anything? How common were they?
Social Psychology	In your country, which forms of prejudice are the most socially acceptable, and which ones are the least acceptable? Why are some forms more acceptable than others?
Think Again	Since inductive arguments are defeasible, how can it ever be reasonable to trust them? Are arguments from analogy really different from inferences to the best explanation?

Some argue that online education is only desirable when face-to-face education is unavailable (Freedman 2013). This paper illustrates the benefits of inverting this proposition: global diversity enables online classrooms to create powerful, previously unavailable educational experiences and new forms of peer education at scale that go “beyond being there” (Hollan and Stornetta 1992).

2 Related Work

A tremendous benefit of diverse classrooms is that students of differing gender, ethnicity, and ability have opportunities to interact. When people interact with similar peers, their shared background leads to automatic thinking. In contrast, interacting with diverse peers often creates a discontinuity (Gurin et al. 2002) that unearths hidden assumptions—yielding more active, effortful and conscious thought (Coser 1975). This active and effortful thinking improves academic performance and makes students more inclusive and democratic (Gurin et al. 2002).

Travel, and interacting with geographically diverse people, similarly induces active thinking and reflection (Lin and Schwartz 2003). For instance, study-abroad programs result in deeper knowledge and understanding—especially about culture and international affairs—and greater self-confidence (Braskamp et al. 2008).

The benefits of interacting with geographically diverse peers arise from differences in experiences and thinking. Examples of these differing experiences include stark differences in population density, income and educational systems (Tudge 2008). People from different parts of the world have different cultural values, reasoning, and preferred learning methods. For instance, cultures differ in their emphasis of individuality versus interdependence (Heine 2008; Markus and

Kitayama 1991) and holistic versus analytical thinking (Varnum et al. 2010). These differences impact cognition. For example, when cultures encourage people to consider objects in relation with their context, they more often apply analogical thinking. By contrast, when people consider objects in isolation, they more often apply categorical rules (Varnum et al. 2010).

To maximize the benefits of diversity, prior work emphasizes two factors: the numeric representation of diverse groups (structural diversity); and the number of settings that students interact in (experiential diversity) (Hurtado et al. 1998b). Ideally, students must meet frequently, and with equal status, in situations where collaboration is necessary and stereotypes are disconfirmed (Pettigrew 1998), and where differing views are welcomed (Hewstone 2003).

Informed by this research, Talkabout forms geographically diverse discussion groups, and encourages fluid roles and consensus-based decisions that emphasize equality. Furthermore, Talkabout contributes a curriculum where students can question stereotypes and compare their views to their peers.

In most current online classes, students' opportunities for discussions with diverse peers are limited to text-based forums. Such asynchronous text channels inhibit trust-formation (Rocco 1998) and open-ended discussion (Short et al. 1976). Synchronous channels, such as video, improve participants' sense of belonging and willingness to collaborate (Saltarelli 2012). Channels such as video which support multimodal communication and nonverbal cues are also better suited to ambiguous discussions (Daft and Lengel 1986) and complex sense-making (DeSanctis et al. 2003). For these reasons, Talkabout leverages synchronous, small-group video discussions to encourage meaningful, open-ended dialogue.

Massive scale presents both a formidable challenge and a powerful opportunity for online education. Prior work encouraging unstructured discussion failed to find an improvement in students' sense of community or academic achievement (Coetzee et al. 2014). More systematically structured approaches have enjoyed greater success. One example is the use of rater redundancy and short exercises that create micro-expertise in peer review: with this structure, peers can provide expert-quality assessment and feedback (Kulkarni et al. 2013), and act as mentors (Papadopoulos et al. 2014). Talkabout introduces a structured interaction and curriculum that leverages diversity.

3 Coordinating Global Small-Group Discussion

The Talkabout interface guides instructors through setting up their course discussions, and creating a structured discussion agenda for students (Fig. 2a). This agenda is displayed throughout the discussion (Fig. 1).

Students choose a discussion time from the published set (Fig. 2b), up to a week in advance. As students log in at their selected time, Talkabout assigns them to groups (instructor can choose group size between 2 and 9). Talkabout has several policies for group assignment; by default it assigns arriving students to a group until

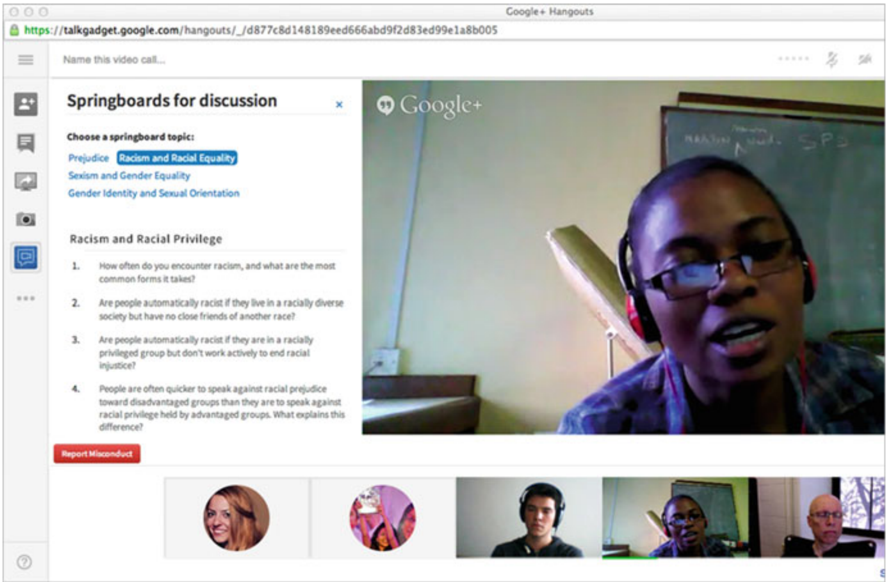


Fig. 1 Talkabout provides a structured discussion agenda and enables students from around the world to discuss with each other

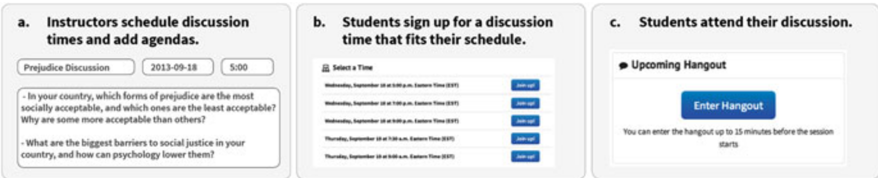


Fig. 2 Talkabout discussion timeline: (**a**) Instructors enter a discussion agenda, and times for the discussion. (**b**) Students pick their preferred time. (**c**) When they log on to Talkabout at their selected time, Talkabout assigns them to a group, and creates a private hangout. Students show up at their selected time, and enter the discussion

it reaches its size limit; then it starts a new group. Other policies, discussed later, explicitly factor geographic location into group assignment. Discussions occur through the Google Hangouts platform for multi-person video and audio chat. For each group, Talkabout creates a discussion session exclusively for the assigned participants. Discussion groups exist only for the duration of the discussion session. If students participate in multiple discussion sessions—even in the same course and on the same topic—they are likely to have different partners, because grouping depends on students' arrival order. Consequently, students hear different ideas and experiences each time.

During discussions, the Talkabout Hangout application shows the instructor's discussion agenda on the left and the video chat on the right. An agenda typically includes suggested discussion topics or activities (Figs. 1 and 4).

4 Assignment by Arrival Yields Diverse Groups

To quantify the geographic diversity in discussions, we aggregate countries into eight geographical regions, and count the number of regions in each discussion. Five regions are from the World Bank's classification (World Bank Group 2012): Eastern Europe and Central Asia (primarily the former Soviet bloc), East Asia and Pacific (mainly China, Japan, Korea, and South-east Asia), South Asia (mainly the Indian subcontinent), Latin America and the Caribbean (Americas except the US and Canada), Middle East and North Africa, and Sub-Saharan Africa. The World Bank only classifies middle- and low-income countries, so we added three other regions: North America (US and Canada), Western Europe, and South Pacific (primarily Australia and Polynesia).

Across seven classes and the first 3200 participants, allocating six-person groups by arrival order yielded discussions with a median of four global regions (Fig. 3b), and a median of five countries (Fig. 3a). The median pair-wise distance between discussants was approx. 6600 km (4100 mi): more than the distance between New York and London.

5 Structuring Talkabout Discussions

Our early experiences with Talkabout, as well as prior work, suggest that it is critical to co-design curricular strategies with educational interaction design. In particular, scripts for discussion have a major impact on student engagement and learning (O'Donnell and Dansereau 1995). Talkabout succeeds best when discussions create opportunities to highlight students' diverse experiences. Based on prior work, we developed three strategies to create discussion scripts or agendas, and refined them through deployments in seven massive classes. Figure 4 shows these

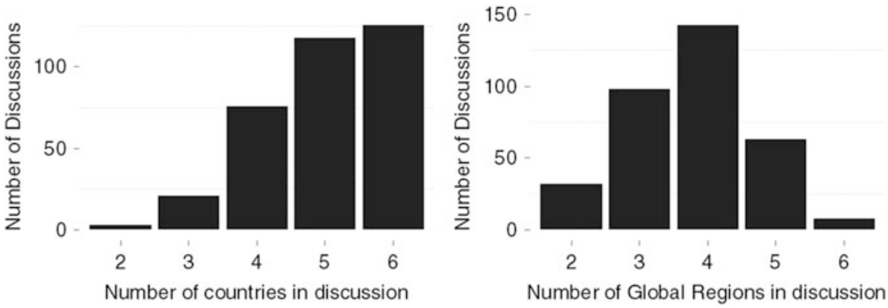


Fig. 3 Across classes. (**a**) Students from many countries participate in each six-person discussion. (**b**) These students aren't just from neighboring countries, they are globally distributed

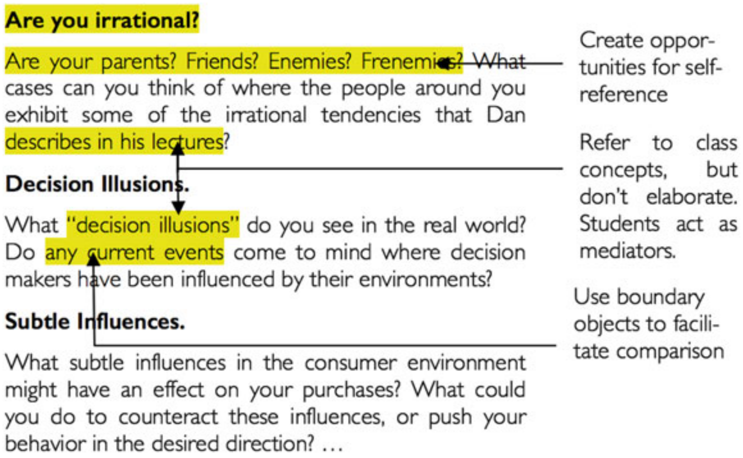


Fig. 4 A representative discussion agenda in a Talkabout, discussion strategies that leverage diversity have been highlighted

strategies embodied in an excerpt from an Irrational Behavior agenda. We discuss each strategy in turn.

5.1 *Create Opportunities for Self-Reference*

Self-reference, when students actively relate class content to their own experiences and perspectives, increases concept elaboration, memory organization, and knowledge retention (Symons and Johnson 1997). Talkabout agendas that employ self-reference ask students to share personal examples that embody class concepts. Self-reference is especially effective when students feel safe in discussing personal experiences. Talkabout groups are small by design to encourage self-disclosure (Nguyen et al. 2012). As each person shares with the group, it encourages peers to likewise disclose (Joinson 2001).

The globally distributed nature of discussions amplifies the benefits of sharing self-referential frames. After a discussion on prejudice in Social Psychology, one student wrote, "I think this may have been the first time the lady from Saudi Arabia had spoken to a Jew [referring to himself]", showing her a different viewpoint. He added, "I told her about the prejudice from Christians I experienced growing up in [US state] in the 1940s and the effect of segregation on blacks," reflecting on his own experience.

Students may see different self-referential frames with different groups. For instance, even though Social Psychology had only one Talkabout discussion (with multiple slots), 454 out of 2553 participants in the Social Psychology class voluntarily attended multiple timeslots.

5.2 Highlight Viewpoint Differences Using Boundary Objects

Talkabout prompts aim to make the differences between students' perspectives salient. This encourages additional self-reference and re-evaluation of previously held theories, which in turn leads to deeper understanding (Guzzetti et al. 1993).

To highlight differences, Talkabout discussion agendas call out boundary objects across geographical contexts. Boundary objects are objects or concepts that maintain their integrity across communities, and yet can be interpreted differently in different communities (Star and Griesemer 1989). Everyday concepts, such as governments, companies/organizations or current events can serve as boundary objects. For instance, one student noted how discussing a 'recent event' yielded new perspective: "we were . . . joined by [a] Syrian. She provided. . .insight of the situation in Syria and how the media is exaggerating it. . . and how the society was quite liberal on Islamic practices (such as wearing the hijab)."

5.3 Leverage Students as Elaborators and Mediators

When a prompt says less, students sometimes say more. Rather than reviewing every relevant concept, Talkabout discussion agendas reference concepts from class without any reminders of what they mean. These underspecified references lead students who have learned these concepts to elaborate, and to act as mediators with students who would have otherwise not understood them. This is similar to highly effective offline strategies like jigsaw classrooms, which also rely on peer-mediated learning and contact with dissimilar peers (Aronson and Bridgeman 2004).

Creating opportunities for mediation also encourages students to ask about other class concepts they haven't understood. For instance, the Organizational Analysis class used "white flight" (a large-scale migration of white Americans to suburbs in the 1950s) as an example of an organizational problem faced by cities. In one Talkabout discussion session, we observed an American student translate the key ideas in this example to a European classmate by making an analogy to intra-European migration.

6 The Anatomy of a Talkabout Discussion

What is the nature of a Talkabout discussion session? We observed and recorded 12 Talkabout discussion sessions in Organizational Analysis. An abridged transcript from an Organizational Analysis class is in Supplementary Materials. Talkabout discussion sessions followed a pattern with clear roles and norms.

6.1 Discussions Follow a Distinct Conversational Pattern

Talkabout discussion sessions usually began with introductions. Since none of the participants knew each other, introductions were fairly formal and detailed. Participants typically shared their first name, their country of residence, and a brief description of their job. Because some participants arrived late to their session, this introduction phase was often repeated.

During these introductions, an informal moderator usually emerged. Moderators often had experience with video-conferencing and a high-bandwidth connection. They exhibited leadership behaviors such as asking participants to introduce themselves, or even explicitly asking to moderate the conversation (e.g. "Shall I lead the conversation?")

After introductions, the informal moderator drew the group's attention to the instructor-provided discussion agenda. Even though agendas sometimes suggested a particular discussion order, participants did not follow it exactly. Instead, they would interpret the agenda for the major theme it embodied, and negotiate what they discussed first. Once students finished discussing a particular prompt, they returned to the agenda to decide the next topic.

While Talkabout discussion sessions were designed to last 30 min, the median length of the discussion was 58 min (Fig. 5). With these longer discussions, students discussed topics that were marked optional, or chose to discuss two topics when the agenda asked only one etc. Many groups also spoke about the class in general after the assigned topics. Conversations typically ended soon after the informal moderator (or a talkative speaker) left the discussion, or when no one in the group suggested a topic to discuss next. As they left, participants often shared how they enjoyed talking to the group, or taking the class. Moderators sometimes encouraged the group to stay in touch after the discussion (e.g. "With the other hangouts, we all added each other on LinkedIn. . . I've already added [name]. If you'd like, feel free to add me.")

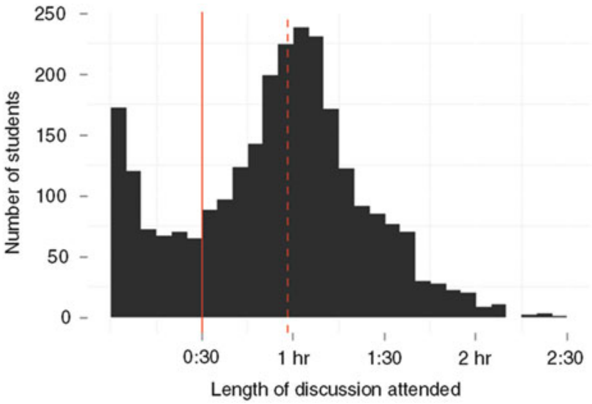


Fig. 5 Across classes, students participated in discussions much longer than instructions indicated. The *solid red line* is the recommended duration for discussion (30 min), the *dashed line* is the median discussion time (58 min)

6.2 Speakers and Spectators

Students seemed to decide early on whether they primarily wanted to speak during the discussion ("speakers"), or listen to the discussion ("spectators"). Spectators often signaled their intent by muting their micro-phones (this showed a "mic muted" icon to others in the discussion).

Speakers tended to be native English speakers or have faster Internet connections. Their discussion was conversational, with overlapping turns similar to face-to-face conversation. Spectators spoke less frequently with longer non-overlapping turns, but were not passive participants. When spectators had trouble finding the right words (e.g., if they were non-native speakers), speakers often suggested words, or encouraged them to continue.

Participants with low-bandwidth connections generally assumed the spectator role and often used the text chat feature in the Google Hangout to "speak" in the discussion. Speakers (usually the moderator) would notice the text, and speak it aloud to the other participants. Both speakers and spectators used text-chat to demonstrate active listening without interrupting the speaker via audio. For example, a student wrote, "Working in [company] must be really cool. Thanks for sharing :)".

A shared video channel forces a single conversation. Still, students sometimes used text-chat as a way for non-discussion related talk, such as exchanging contact information or LinkedIn profiles.

7 Study 1: Do Discussions Improve Performance?

It is not obvious that the benefits of peer discussions (Brookfield and Preskill 2012; Parker 2000) would transfer to an online environment. In these environments, peers have vastly different backgrounds and no prior interaction with each other. Therefore, our first study measures the benefits of participation in online discussions. Later experiments measure how these benefits vary with geographic diversity in discussion groups.

With many educational practices, it is difficult to draw a causal link between participation and student learning. For instance, students may self-select to participate. To combat this bias, we use a control condition in which interested students are actively prevented from discussing. Furthermore, we use an intention-to-treat analysis that recognizes that some students will not participate, even when given the opportunity. Therefore, this analysis asks: after controlling for students that don't discuss given an opportunity, are discussions effective? Such analysis is common in clinical trials, where patients that are randomly assigned to a treatment group are included in the analysis even if they do not take their medication. Because intention-to-treat analyses take non-compliance into account, they result in conservative estimates of a drug's effectiveness.

7.1 Method: Wait-List Control

In a between-subjects experiment, we randomly assigned students in the Organizational Analysis class on Coursera to either a Discussion condition, or to a Wait-list condition. This assignment occurred when they signed up for a discussion time on Talkabout, after consenting to participate in the study.

Students in the Discussion condition were allowed to participate in discussions starting in Week 1, while those on the wait-list were not allowed to participate in discussions until Week 5. This setup results in two discussion opportunities (Weeks 1 and 3) where a subset of students was prevented from participating. Even though some participants in the Discussion condition did not attend discussion, they were included in the intention-to-treat analysis.

7.2 Hypotheses and Measures

We hypothesized that participating in a Talkabout discussion session would motivate students to engage with other course components. Prior work similarly finds that discussions motivate students to engage with in-person classes (Brookfield and Preskill 2012). To measure engagement, we check whether the student participated in the course quiz due the day after discussion. Recall that participation in MOOCs is entirely voluntary, and several classes have battled with attrition (Ho et al. 2014). Quizzes are a high-effort activity that most MOOC learners don't participate in: only 22.8 % of students who watched a lecture video also participated in a quiz. This makes quizzes suitable as a high-effort engagement measure (Coetzee et al. 2014; Tomkin and Charlevoix 2014).

We further hypothesized that students in the Discussion condition would do better on the quiz, aided by the self-reference, reflection and revision of class concepts.

7.3 Participants

Overall, 1002 students were assigned to the Discussion condition, and 122 to the Wait-list condition. We used an unbalanced design to maximize the number of students who benefited from discussions. Of those in the discussion condition, 397 attended a discussion.

7.4 Results: Discussion Increases Class Participation, Marginally Improves Grades

Students in the Discussion condition were more likely to take the quiz. A logistic regression indicated that odds of taking the quiz were 1.46 times higher for the Discussion condition (Wald $z^* = 1.97$, $p < 0.05$). Students in the Discussion condition also did marginally better on the quiz ($t(1122) = 1.89$, $p = 0.06$).[1] The average improvement was 16.7 %.

Thus, even accounting for students who do not follow through, discussions help students stay engaged in the course and perform better on related assessments.

While Talkabout participation improves engagement, this effect seems short-lived. Students who participate in a Talkabout 1 week are not more likely to participate in the quiz the following week: Wald $z^* = 1.61$, $p = 0.10$. We also found no significant improvement in quiz scores for the quiz due the following week.

Would participating in multiple Talkabout discussion sessions improve these short-term benefits? As is typical with online classes, many students shopped the first weeks, and only 113 students in the discussion condition attended the second discussion (397 attended the first week). Therefore, our intention-to-treat analysis lacks the statistical power to capture any benefits of participating in multiple discussions. Also, while the wait-list design can control for intent to participate, students who actually participate in discussions may still differ from those who don't (e.g. they could be more motivated). An intention-to-treat analysis estimates effects by assuming participants' distribution (e.g., for motivation) are similar in the wait-list and treatment groups due to randomized assignment, but this experiment does not verify this assumption.

The results of this study suggest that performance on class quizzes may improve even with limited participation, and that discussions improve student engagement. Do these effects depend on the participants in the discussion? Given our hypothesis that geographic diversity should help learning, our next study investigates the effect of discussants' geographic diversity on course performance.

8 Study 2: Does Diversity Improve Discussion Benefits?

Study 1 established that participating in Talkabout discussions improves class engagement. Is geographic diversity causing this effect? In a second, between-subjects experiment, Talkabout's group-assignment algorithm randomly assigned students either to a single-region group or a multi-region group. Participants regions were determined by the five World Bank regions, as well as three regions to capture

[1] While only marginally significant ($p < 0.10$), we include this result because it is suggests opportunities for future work.

North America, Western Europe and the South Pacific. The Same-region condition grouped students with others from their region. The Multi-region condition grouped students from anywhere in the world. We discarded data from the South Pacific region because it had few participants.

8.1 Participants and Setup

Fifty-five students in the Organizational Analysis class participated. When students logged on to the site, we recorded their IP address, found their location based on IP, and randomly assigned them to the Multi-region high-diversity or the Same-region low-diversity condition. Students were then grouped into discussion groups with a maximum of six participants.

8.2 Measures

To measure conceptual understanding, we invited students to fill out a questionnaire immediately after the discussion; 43 participated. We asked students to answer to the best of their ability, but informed them that their answer would not affect their course grade. This survey had one open-ended question which required critical thinking and an under-standing of concepts discussed in the session ("Where would you want to position yourself if you wanted leverage over the flow of "information" in a social network—centrally, peripherally, or in a bridging position. Why?"). We scored this question in consultation with the teaching assistant of the course. The average score was 47 % (combining both conditions). We use students' grade in a prior class quiz as a measure of prior performance (we ignore data from one participant, who did not complete the quiz). The questionnaire also asked questions about how much they liked their discussion, and how much they felt they learned from it.

8.3 Hypothesis

Students in the Multi-Region, high-diversity, condition were exposed to more contrasting viewpoints and self-reference than discussions in low-diversity groups. Thus, we hypothesized that members of more geographically diverse groups would have higher scores on the post-questionnaire.

8.4 Manipulation Check

The median number of countries in the same-region condition was two (both from the same geographical region), while the median in the multiple-region condition was four.

Does large geographical distance imply a diverse group? Some World Bank regions are large, so we examined if multi-region groups had more differing national viewpoints than same-region groups, taking into account how economic opportunities and educational experience influence everyday experience (Desai 1991), as do cultural values (Markus and Kitayama 1991).

We used each participant's country to map them onto diversity attributes used in cultural psychology and political science. We use countries as our unit of analysis because they have a consistent typology of collectivistic or individualistic culture (Green 2005), organizational attitudes such as inter-personal dependence and criteria for fulfillment (Shenkar and Oded 1985), economic development (Goesling 2001) and life expectancy (Marmot 2005). While each country is diverse, within-country differences are smaller than between-country differences (Green 2005), making this by-country analysis feasible.

We compared countries of participating students on three dimensions: cultural values, income, and pupil-teacher ratios in primary school. As a measure of cultural values, we used the mean overall secular values for each country from the World Values Survey (2014). Countries with lower scores have societies that emphasize religion, traditional family values, and collectivistic thinking. The average pair-wise difference between participants' countries on the overall secular values scale was lower in the same-region condition than in the multi-region condition, Wilcoxon $W = 407.5$, $p < 0.05$ (same-region mean: 0.022, equivalent to the difference between the US and Romania, multi-region mean: 0.031, equivalent difference: US and Thailand).

Students' countries in the Multi-region condition had marginally higher differences in income levels compared to those in the Same-region condition ($t(74) = 1.81$, $p = 0.07$; log-transformed because income distribution is log-normal (Clementi and Gallegati 2005)). Using data from the World Bank (World Bank Group 2012), the median per-capita annual income differed on average by \$8120 (PPP) in the same-region condition, approximately the difference between the US and Canada. The average difference in the multi-region condition was \$20,495 (PPP), approximately the difference between the US and Israel.

Lastly, students' countries in the multi-region condition had greater pairwise variation in educational experience, as reflected in primary school pupil-teacher ratios ($t(74) = 2.00$, $p < 0.05$). Using World Bank data (World Bank Group 2012), the median differences in the pupil-teacher ratios in the same-region condition were 2.91 (approximately the difference between schools in the US and Canada), while the median difference in the multi-group condition was 5.91 (the difference in schools between the US and Russia).

Collectively, these analyses suggest that multi-region groups brought more diverse experiences and backgrounds to their discussions.

8.5 Results: Students in Diverse Groups Perform Better

On a 7-point Likert scale question, students in the high-diversity condition rated their discussion as more enjoyable than those in low-diversity (Mann–Whitney $U = 140.5$, $p < 0.05$). They also reported learning marginally more from their discussion partners on a different 7-point Likert scale (Mann–Whitney $U = 160.5$, $p = 0.08$).

Based on the grades in the post-quiz, an ordinary-least-squares linear model showed that after controlling for prior performance, students in the high-diversity condition out-performed those in the low-diversity condition, ($\beta = 0.41$, $F(1,37) = 2.31$, $p < 0.05$, adjusted $R2 = 0.11$). A post-hoc comparison also found that students in discussions with more countries did better in both conditions. Using an ordinary-least-squares linear model, we found that the number of countries in the discussion was predictive of the quiz score ($\beta = 0.15$, $F(1,36) = 2.57$, $p < 0.01$, adjusted $R2 = 0.14$).

This result suggests that even countries in the same geo-graphical region add meaningful diversity. This may be because regions are too large and diverse (e.g. the Latin America and Caribbean region has 35 countries). There-fore, counting countries rather than regions may provide a better measure of diversity.

However, this experiment only measures the immediate effects of diversity in a single class. Do geographically diverse discussions have a longer-term effect, and do these benefits generalize across classes? We now describe a longitudinal deployment that evaluates the effect of di-verse discussions on grades in actual course tests over periods of weeks.

9 Study 3: Large-Scale Field Experiment

In Study 3, we sought to confirm and expand upon Study 2's diversity effect across more classes and with more students. In doing so, we trade off some of Study 2's experimental control in exchange for a much larger sample. We conducted our experiment across two large online classes, Organizational Analysis and Social Psychology.

9.1 Participants

In the Social Psychology class, 2025 students participated. In the Organizational Analysis, 397 students participated.

All students in the Organizational Analysis class who wanted to participate in discussions used Talkabout. By the instructor's request, the Social Psychology class also al-lowed students to choose an in-person discussion instead. In-person discussants received the same discussion agenda and directions as online discussants. Two thousand and thirty-seven students reported participating in an in-person discussion. Except for qualitative comparisons between online and in-person discussions, we ignore their data. It is possible that online discussions attracted students who believed they would benefit more from a diverse discussion. However, the main results of this study were consistent across both classes.

9.2 Method

Similar to Study 2, Talkabout grouped students into discussions. However, students were not explicitly grouped into high- and low-diversity conditions. Instead, this study used a simpler approach where Talkabout collected participants in order of arrival. When a group had six students, Talk-about launched a new group. This setup assigns participants to diversity levels in a random fashion. Participants in both classes had no control over who their discussion partners were, and therefore had no control over the level of geographic diversity in their discussion.

The two classes implemented different schedules for their discussions. Social Psychology held discussions for 1 week at the end of class, 2 weeks before the final exam. Organizational Analysis had discussions throughout the class, starting from the first week. This variety allows us to understand the effect of Talkabout both for highly motivated students who remain active at the end of class, and for enthusiastic, but potentially uncommitted learners.

9.3 Hypotheses and Measures

We hypothesized that participating in more geographically diverse Talkabout discussions would lead to better course performance, as students became more active thinkers through conversations with diverse students. In addition, given our results in Study 1, we hypothesized that students in more diverse discussions early in the class would stay engaged with the class for longer.

To measure geographic diversity, we use the number of countries in a discussion as a coarse but useful metric. While students using Talkabout may be systematically different from the median resident of their country (they can afford an Internet connection), national cultures still importantly shape their thoughts and actions (Hambrick et al. 1998).

To measure performance, in Social Psychology, we used the final exam score. The final exam was a 50 multiple-choice question test (see Appendix for a sample of questions). The instructor created this exam independently with no input from the research team. The Organizational Analysis class had weekly quizzes due every Sunday, which we use as a performance measure. The instructor independently created these quizzes in a previous run of the class (before Talkabout was designed), and they were used unchanged in the experimental class. The first Talkabout session was 1 day before the first quiz was due. We analyze the first 2 weeks' quizzes. The first quiz had 19 multiple-choice questions; the second had 16 (see Appendix 2). Finally, both classes invited students to participate in a post-discussion survey about their experience.

9.3.1 Analysis Procedure

For both classes, we built an ordinary-least-square linear regression for performance based on the number of countries in the discussion. Because the number of discussants and number of countries is collinear ($R2 = 0.81$ and 0.88 in the two classes), we only analyzed groups of six students. We controlled for each student's prior performance in class if any previous quizzes had occurred. Our model for the first week's quiz in Organizational Analysis had no measure for prior performance (model $R2 = 0.003$). The model for the second quiz ($R2 = 0.11$) used the score in the first quiz as a prior-performance metric. The model for the Social Psychology class ($R2 = 0.05$) used a student's total grade in all assignments before the final exam as a prior-performance metric.

9.4 Results

Our analysis finds support for the first hypothesis: students perform better on tests after a more geographically diverse discussion. We find no support for our second hypothesis that diverse discussion improves retention in the long-term.

9.4.1 High-Diversity Discussions Improve Scores

In both classes, more diverse discussions led to higher exam grades (Table 2). In Social Psychology, on the final exam out of 50 points, each additional country adds an approximate $\beta = 1.78$ points (2.4 % of the final grade) to a student's final exam score ($t(129) = 1.78$, $p = 0.01$). In Organizational Analysis, on the Week 2 quiz out of 16 points, each additional country yields $\beta = 0.39$ points (3.6 %) to the quiz score ($t(110) = 2.03$, $p < 0.05$). However, from the model for the Week 1 quiz (without a prior-performance measure), we do not see any significant effect of diversity on score. Prior performance helps capture sufficient variation to make diversity statistically distinguishable from a null hypothesis.

Table 2 After controlling for prior performance, more countries in a discussion lead to better grades, in both Social Psychology and Organizational Analysis

	β	F	p-Value
Organizational Analysis: Week 1 Quiz ($R^2 = 0.003$)			
Intercept	15.7	21.51	<0.001
Number of countries	0.11	0.76	0.46
Organizational Analysis: Week 2 Quiz ($R^2 = 0.11$)			
Intercept	8.11	4.33	<0.001
Week 1 grade (z-scored)	0.78	2.81	<0.001
Number of countries	0.39	2.03	0.02
Social Psychology: Final Examination ($R^2 = 0.05$)			
Intercept	27.20	7.00	<0.001
Pre-final grade (z-scored)	0.91	1.30	0.19
Number of countries	1.78	2.34	0.01

9.4.2 Benefits of Diverse Discussions Last Roughly 2 Weeks

In the Organizational Analysis class, while geographic diversity leads to better quiz scores 1 week after discussion (Week 2 quiz), we did not find any significant effects into Week 3. Similarly, we built an ordinary-least-squares linear model for predicting how many weekly quizzes a student would participate in, based on the number of countries in their first discussion. We found no significant effect ($t(130) = -0.49$, $R2 < 0.001$). Similar to results from Study 2, this suggests that the benefits of a diverse discussion only persist for a short duration.

9.4.3 Geographic Diversity Leads to New Perspectives

Post-discussion, a survey asked participants about the best part of their discussion. Two independent raters coded 100 responses about whether comments mentioned participant diversity: 51 % mentioned it (Cohen's $\kappa = 0.7$, $z = 7.04$, $p < 0.001$). Students noted that diversity yielded different experiences and examples and perspectives, which challenged ones held by students. A Social Psychology student wrote how they learned that "…in China it is a custom for married women to keep their surnames, thus I [now] think women changing their surnames when married in other countries has something to do with sexism." An Organizational Analysis student said, "It was interesting to hear about organizations in Australia, Ukraine, Israel, Indonesia, and Canada. Similar issues appear everywhere regarding decision-making."

9.4.4 Gender Representation Does Not Influence Scores

In prior work, the proportion of females participants affected collaborative group outcomes (Woolley et al. 2010). However, in our study, female participation did not affect performance after controlling for the number of countries in each group. Adding the proportion of female participants to the Organizational Analysis class model for the Week 2 quiz did not improve model fit, and the effect of gender was not significant: $t(100) = 1.1$, $p = 0.26$. The Social Psychology class shows a similar non-significant effect: $t(128) = 0.62$, $p = 0.53$.

9.4.5 Other Non-significant Factors

We test the following variables in isolation; all were non-significant with $p > 0.50$. We found no significant effect of the arrival order of participants on either the diversity in their group, or the benefits of diversity on course grades. We also find no evidence that diverse discussions had larger benefits for either gender. Finally, there was no significant correlation between how early students signed up for a discussion and their benefits.

9.4.6 Other Measures of Geographic Diversity

The results of our analysis were consistent when we used other measures such as the pairwise distance between participants' locations. We use the number of countries while describing results because it is more interpretable.

9.5 Limitations

This experiment included two classes, Social Psychology and Organizational Analysis. Both classes used Talkabout in discussions focused on critical thinking and sense-making. As such, evidence that geographically diverse discussions improve engagement and learning may not generalize to classes that emphasize procedural knowledge (e.g. Corporate Finance), or classes where benefits from global perspectives are smaller (e.g. physics). That said, even the most procedural topics require critical thinking and judgment, and as many instructors have found, topics like physics that seemingly don't benefit from global perspectives may still benefit from discussions (Crouch and Mazur 2001; Mazur 2009).

Geographic diversity encodes many other kinds of diversity, e.g., economic opportunities, cultural values, and education experience. Each of these dimensions may have differing benefits for online classes. Future work could build theory that differences matter when.

10 Discussion

It can be difficult to demonstrate measurable learning effects using design interventions in online courses. For example, while it is possible to increase student involvement in forums (Anderson et al. 2014), improving grades and retention has remained challenging (Coetzee et al. 2014; Tomkin and Charlevoix 2014). However, Talkabout increases both learning and engagement (Table 3). One reason for this improvement may be that Talkabout developed a pedagogical approach alongside the software. In pilots without meaningfully structured discussions, it fared poorly. Furthermore, Talkabout builds a social environment and an opportunity for reflection. It does this via a medium that is known to build trust (Rocco 1998) and is suited for open-ended discussions (Short et al. 1976), such as those leading to sense-making (DeSanctis et al. 2003).

Geographic diversity's direct effect is in students meeting people from other world regions. It is associated with changes in several other diversity measures (e.g., cultural values, economic opportunity, and educational experience). This paper demonstrates that geographic diversity indeed impacts these other measures. However, there may be other causal pathways involved. It is possible that students who differ in geographical location still have similar socio-economic backgrounds, and students who live very close may be very different. Future work can develop more nuanced diverse experiences.

Talkabout also points to the benefits of using video for geographically diverse discussions. Video conferencing creates a middle ground of immersion in another culture. With complete immersion in an in-person setting, the norms and views of the majority are pervasive (Hurtado et al. 1998a; Nora and Cabrera 1995). Students with a minority viewpoint in a fully-immersive experience may find themselves confronted with the choice to either embrace the majority culture (suppressing their own), or reject it and flounder (Ogbu 1992). On the other hand, with the minimal immersion, say, of lectures, students may ignore alternate viewpoints as a mere academic exercise. Video-conferencing may occupy an attractive middle ground: it is interactive, compelling students to engage with their diverse classmates and reflect upon their contact (Lin and Schwartz 2003). One student told us in an interview, "Talkabout helps bring the class together—it makes the learning tangible and real. . .you are interacting with other people, who are experiencing a lot of different things."

Table 3 Summary of experimental results

Study 1: Discussion participation with a wait-list control
Participating in a video discussion with peers increases participation in quizzes and marginally improves performance.
Study 2: Controlled manipulation of geographic diversity
Students in high geographic diversity discussion groups perform higher.
Study 3: Large-scale study of geographic diversity
High geographic diversity discussions lead to improved short-term performance in two classes, but do not improve multi-week retention.

Video-based discussions are not without their problems today. Some countries (e.g., Iran) restrict access to Google Hangouts, low-bandwidth connections degrade the student experience, and installing video-conferencing software remains challenging for some students. However, these technological limitations are likely to lessen as bandwidth becomes more plentiful and software comes pre-installed.

10.1 Comparing In-Person and Online Discussions

Recall that Social Psychology allowed students to choose to run their discussion in person instead of online. Students participating in the in-person discussions often turned to close friends and relatives. The shared context made the conversation friendlier. For instance, one participant re-marked, "I really like the discussion because it was with my friends… It was really easy to start the discussion." In-person discussions also had lower geographic diversity. One student summarized, "Being from the same age group, social level and from the same community; we had very much similar views about the topics in hand." Students reported difficulties scheduling discussions and keeping them on-topic. One remarked, "We had to reschedule a couple of times [before we could meet]." And with friends, "Turning a conversation towards a scientific discipline such as social psychology was hard and a bit artificial…" An-other remarked, "Members were my family… and speaking about some things is not easy!"

10.2 The Design Space of Online Peer Conversations

Talkabout currently implements a particular design for online discussions. To arrive at this design, we explored a number of different decisions in this design space (Table 4).

10.2.1 Always-Available Discussions Lack Critical Mass

Always-available and unscheduled discussions in classes may enable students to talk with a remote partner whenever they have a question or thought. To test the feasibility of this idea, we created a version of Talkabout where students could sign up for an immediate discussion. If another student indicated their availability within the next hour, Talkabout would email both to set up a discussion.

We tested this version in the Think Again philosophy and argumentation class over a three-day period. Of the 2940 who saw the opportunity, 54 students signed up. Unfortunately, only five students overlapped within the 1-h window. This suggests that MOOCs attract many students, but their presence on the course site does not spontaneously overlap. Therefore, Talkabout instead adopts a bus stop model where discussions occur at regular time intervals, making critical mass more likely.

Table 4 Talkabout's current implementation emphasized in italics, design choices that we found to be worse are emphasized in bold

Design dimension	Choices		
Same discussants every time?	*Yes*	**No**	When possible
Group size	**Small**		Large
Discussion guidance	None	**Guidelines and prompts**	*Scripts*
Role negotiation	Instructor specifies	*Technologically mediated*	**Informal**
Discussion scheduling	*On-demand any time*	**Bus stop: regular intervals**	*Same time every week*

10.2.2 Students Prefer to Negotiate Roles Informally

Prior work suggests including a designated discussion facilitator to attend to group dynamics in distributed discussions (Mark et al. 1999). Could formal facilitators improve Talkabout discussions? We conducted a between-subjects experiment with two conditions ($n = 80$) in the Organizational Analysis class. In the facilitator condition, all participants in a Hangout saw a button to volunteer to be a discussion facilitator. When a student volunteered, the system would show them facilitation tips. Other participants saw a message that the volunteer was facilitating the discussion. In the control condition, students were not shown the button to volunteer. Of the 40 students in the facilitator condition, seven volunteered. An intention-to-treat analysis showed a trend to-ward students in the facilitation condition feeling the discussion was less motivating (Mann–Whitney $U = 191.5$, $p = 0.09$), and a trend toward less willingness to meet the same group again ($U = 191.5$, $p = 0.09$). These results suggest that fluid negotiation of moderation may work better than a formal facilitation role.

10.2.3 Rigidly Enforced Scripts Lower Satisfaction

Prior work in CSCL suggests that structuring collaboration between students using instructions or scripts yields improved learning (O'Donnell and Dansereau 1995). What is the right degree of scripting? In a between-subjects experiment ($n = 82$) in the Organizational Analysis class, we explored the benefit of an enforced script. In this condition, Talkabout only showed the current discussion topic, and participants needed to click a button to indicate completion and advance to the next topic. The control condition agenda showed all topics at once.

Of the 50 students in the enforced-script condition, only four clicked the "next topic" button even once. In the post-survey, students also reported they felt the discussion was less motivating (Mann–Whitney $U = 193$, $p = 0.07$), and that they were less willing to meet the same group again ($U = 191.5$, $p = 0.08$). This suggests that enforcing a discussion order may undermine the social benefits of Talkabout (Dillenbourg 2002).

10.2.4 Same-Partner Discussions Have Inadequate Participation

In the in-person classroom, it is common practice to assign students to groups with fixed membership for the duration of a project or series of discussions throughout a course (Brookfield and Preskill 2012). Repeated interactions in such groups build trust and rapport (Bos et al. 2002). By contrast, non-persistent groups lack familiarity but expose students to different viewpoints.

In a between-subjects experiment in the Think Again class ($n = 522$), we randomly assigned students to either a persistent or control condition. The persistent condition assigned students to the same group for every discussion. The control condition assigned them to a group when they arrived to the site, as described previously. Students in both conditions attended the same number of discussions ($\mu = 0.46$, $t(522) = 0.33$, $p = 0.73$). However, as students dropped the class, the size of discussion groups in the persistent condition kept shrinking until they were no longer viable. While 27 % of the control groups had at least 5 discussants, only 2 % of the persistent groups did ($t(81) = 4.67$, $p < 0.001$). Therefore, our discussion strategies structure discussions to leverage changing partners. The next step might be to forge a middle ground where Talkabout prefers familiar partners but adapts groups if previous partners drop out.

11 Conclusion

This paper suggests that the geographic diversity in online classes can be an educational asset. Instead of becoming a handicap, distance can expose students to others and to other ways of thinking. However, leveraging the diversity of online environments requires careful design. This paper describes one such approach, Talkabout, which uses video chat to create discussions between students across the world. Embracing and designing for diversity can enable other innovations. For instance, instructors could leverage students as co-creators and draw on students' local observations to showcase how course concepts arise differently around the world. Likewise, international relations or security courses might launch a global crisis simulation with each student representing their own region. These educational experiences offer a glimpse of the potential of thinking "beyond being there" (Hollan and Stornetta 1992). They are not just leveraging geographic diversity—they would be impossible without it.

Acknowledgments We are grateful to Dan Schwartz for thought-provoking discussions and feedback with early drafts; Coursera and the OpenEdX teams for platform integration and encouraging instructors to try Talkabout; instructors of the many MOOCs who use Talkabout in their classes, and the many students who connected with their peers and shared their stories with us. This research was funded in part through NSF grants #1351131 and #1444865, the Hasso Plattner Design Thinking Program, and the Siebel Scholars Program. This research was conducted under Stanford IRB protocol #30319.

References

Anderson A et al (2014) Engaging with massive online courses. In: Proceedings of the 23rd international conference on world wide web, ACM

Aronson E, Bridgeman D (2004) Jigsaw groups and the desegregated classroom: in pursuit of common goals. In: Aronson E (ed) Readings about the social animal. Worth Publishers, New York, NY, p 532

Bos N, et al. (2002) Effects of four computer-mediated communications channels on trust development. In: Proceedings of the CHI: ACM conference on human factors in computing systems. ACM Press

Braskamp LA et al (2008) Assessing progress in global learning and development of students with education abroad experiences. Front Interdiscip J Study Abroad 18:101–118

Breslow LB et al (2013) Studying learning in the worldwide classroom: research into edX's first MOOC. Res Pract Assess 8:13–25

Brookfield SD, Preskill S (2012) Discussion as a way of teaching: tools and techniques for democratic classrooms. Wiley, San Francisco

Clementi F, Gallegati M (2005) Econophysics of wealth distributions. In: Chatterjee A et al (eds) Econophysics of wealth distributions. Springer Milan, Milano

Coetzee D et al (2014) Chatrooms in MOOCs: all talk and no action. In: Proceedings of the ACM conference on Learning @ scale. ACM Press, pp 127–136

Coser RL (1975) The complexity of roles as a seedbed of individual autonomy. In: Coser LA (ed) The idea of social structure: essays in Honor of Robert Merton. Harcourt Brace Jovanovich, New York

Crouch CH, Mazur E (2001) Peer instruction: ten years of experience and results. Am J Phys 69(9): 970

Daft RL, Lengel RH (1986) Organizational information requirements, media richness and structural design. Manage Sci 32(5):554–571

Desai M (1991) Human development: concepts and measurement. Eur Econ Rev 35(2–3):350–357

DeSanctis G et al (2003) Learning in online forums. Eur Manage J 21(5):565–577

Dillenbourg P (2002) Over-scripting CSCL: the risks of blending collaborative learning with instructional design. In: Kirschner PA (ed) Three worlds of CSCL. Can we support CSCL? Open Universiteit Nederland, Heerlen, pp 61–91

Freedman J (2013) MOOCs are usefully middlebrow. The chronicle of higher education. http://chronicle.com/article/MOOCs-Are-Usefully-Middlebrow/143183/

Goesling B (2001) Changing income inequalities within and between nations: new evidence. Am Sociol Rev 66(5):745–761

Green EGT (2005) Variation of individualism and collectivism within and between 20 countries: a typological analysis. J Cross-Cult Psychol 36(3):321–339

Gurin P et al (2002) Diversity and higher education: theory and impact on educational outcomes. Harv Educ Rev 72(3):330–367

Guzzetti BJ et al (1993) Promoting conceptual change in science: a comparative meta-analysis of instructional interventions from reading education and science education. Read Res Q 28(2): 116–159

Hambrick DC et al (1998) When groups consist of multiple nationalities: towards a new understanding of the implications. Organ Stud 19(2):181–205

Heine SJ (2008) Cultural psychology. WW Norton, New York
Hewstone M (2003) Intergroup contact: panacea for prejudice? Psychologist: 352–355. http://www.psy.ox.ac.uk/publications/28661
Ho AD, Reich J, Nesterko S, Seaton DT, Mullaney T, Waldo J, Chuang I (2014) HarvardX and MITx: the first year of open online courses. HarvardX and MITx Working Paper No. 1
Hollan J, Stornetta S (1992) Beyond being there. In: Proceedings of the SIGCHI conference on human factors in computing systems—CHI'92. ACM Press, pp 119–125
Hurtado S et al (1998a) The climate for diversity: key issues for institutional self-study. New Dir Inst Res 1998(98):53–63
Hurtado S et al (1998b) Enacting diverse learning environments: improving the climate for racial/ethnic diversity in higher education. ASHE-ERIC Higher education report, Vol. 26, No. 8
Jacobs AJ (2013) Grading the MOOC University. The New York Times
Joinson AN (2001) Knowing me, knowing you: reciprocal self-disclosure in internet-based surveys. CyberPsychol Behav 4:5
Konstan JA et al (2014) Teaching recommender systems at large scale. In: Proceedings of the ACM conference on learning @ scale conference. ACM Press
Kucsera J, Orfield G (2014) New York state's extreme school segregation: inequality, inaction and a damaged future. Educ Rev 79(2):269–284
Kulkarni C et al (2013) Peer and self assessment in massive online classes. ACM Trans Comput Hum Inter 20(6):33
Lin X, Schwartz DL (2003) Reflection at the crossroads of cultures. Mind Cult Act 10(1)
Losh E (2014) The war on learning: gaining ground in the digital university. MIT Press, Cambridge
Mark G et al (1999) Meeting at the desktop: an empirical study of virtually collocated teams. ECSCW. Kluwer Academic Publishers, Dordrecht
Markus H, Kitayama S (1991) Culture and the self: implications for cognition, emotion, and motivation. Psychol Rev 98(2):224–253
Marmot M (2005) Social determinants of health inequalities. Lancet 365(9464):1099–1104
Mazur E (2009) Farewell, lecture? Science 323(5910):50–51
Nemeth CJ (1986) Differential contributions of majority and minority influence. Psychol Rev 93: 23–32
Nguyen M et al (2012) Comparing online and offline self-disclosure: a systematic review. Cyberpsychol Behav Soc Netw 15(2):103–111
Nora A, Cabrera AF (1995) The role of perceptions in prejudice and discrimination and the adjustment of minority students to college. J High Educ 67(2):119–148
O'Donnell AM, Dansereau DF (1995) Scripted cooperation in student dyads: a method for analyzing and enhancing academic learning and performance. In: Hertz-Lazarowitz R, Miller N (eds) Interaction in cooperative groups: the theoretical anatomy of group learning. Cambridge University Press, Cambridge, pp 120–141
Ogbu JU (1992) Understanding cultural diversity and learning. Educ Res 21(8):5–14
Olds K (2013) Mapping Coursera's global footprint. Inside Higher Ed. http://www.insidehighered.com/blogs/globalhighered/mapping-courseras-global-footprint
Papadopoulos K et al (2014) Community TAs scale high-touch learning, provide student-staff brokering, and build esprit de corps. In: Proceedings of the 1st ACM conference on Learning @ scale conference—L@S'14. ACM Press, pp 163–164
Parker WC (2000) Classroom discussion: models for leading seminars and deliberations. Soc Educ 65(2):111–115
Pettigrew TF (1998) Intergroup contact theory. Annu Rev Psychol 49:65–85
Purdie N, Hattie J (1996) Cultural differences in the use of strategies for self-regulated learning. Am Educ Res J 33(4):845–871
Rocco E (1998) Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact. In: Proceedings of CHI: ACM conference on human factors in computing systems. ACM Press

Saltarelli A, Roseth C, Glass C (2012) Effects of belongingness and synchronicity on face-to-face and online cooperative learning. Society for information technology & teacher education international conference, vol 2012(1)

Shenkar S, Oded R (1985) Clustering countries on attitudinal dimensions: a review and synthesis. Acad Manage Rev 10(3):435–454

Short JE et al (1976) The social psychology of telecommunications. Wiley, London

Star SL, Griesemer JR (1989) Institutional ecology, 'Translations' and boundary objects: amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. Soc Stud Sci 19(3):387–420

Symons CS, Johnson BT (1997) The self-reference effect in memory: a meta-analysis. Psychol Bull 121(3):371–394

Tomkin JH, Charlevoix D (2014) Do professors matter? In: Proceedings of the ACM conference on Learning @ scale. ACM Press, pp 71–78

Tudge J (2008) The everyday lives of young children. Cambridge University Press, Cambridge

Varnum MEW, Grossmann I et al (2010) The origin of cultural differences in cognition the social orientation hypothesis. Curr Dir Psychol Sci 19(1):9–13

Woolley AW et al (2010) Evidence for a collective intelligence factor in the performance of human groups. Science 330(6004):686–688

World Bank Group (2012) World Development Indicators 2012. World Bank Publications

World Values Survey (2014) World Values Survey (Wave 6). World Values Survey Association

Improving Design Thinking Through Collaborative Improvisation

David Sirkin, Brian Mok, Stephen Yang, Rohan Maheshwari, and Wendy Ju

Abstract Over the last 2 years, we have been following an improvisational approach to physical interaction design research. It emphasizes the use of exploratory lab and field experiments as a way to (a) source novel ideas about how people might interact with expressive objects such as robots and active spaces, (b) appraise the performance of our prototypes of these technologies, and (c) build frameworks to understand users' mental models and develop new insights into interaction. We have focused, in particular, on staging environments—whether in public settings or recreated in our workspace—where we can provoke discussion about what behaviors and emotions would be desirable or natural. This paper describes how we design and run experiments to evaluate how people interact with expressive robots built from everyday objects, including a mechanical ottoman, emotive dresser drawers and roving trash barrel.

1 Introduction

> Improv. . .allows the designer to explore the design solution in all sensory and cognitive modalities. . .in a way that cannot be achieved through mere graphic design or well-plotted "user scenarios." While the actor uses empathy to perform dramatic characters in scripted situations, the designer uses empathy to perform design solutions that are drawn from deep identification with real, individual people in specific situated contexts in the real world.
>
> *-Brenda Laurel*
>
> *Design Research: Methods and Perspectives* Laurel (2003)

Our current work explores, how design engineers can draw out people's implicit intuitions and expectations for how interactions with machines should transpire. In this paper, we focus on our use of semi-structured, improvised interaction sessions in order to prototype devices and platforms, experiment with prompts and responses, and test social norms.

D. Sirkin (✉) • B. Mok • S. Yang • R. Maheshwari • W. Ju
Center for Design Research, Stanford University, 424 Panama Mall, Stanford, CA 94305, USA
e-mail: sirkin@stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_7

The approach evolved from our research in physical interaction design and robot mediated communication: domains that rely on designers' knowledge and intuition of embodied action, which are often only implicitly understood, and therefore can be difficult to verbalize. Our use of collaborative improvisation has therefore been particularly helpful in revealing, and making explicit, these internal thought processes.

2 Approach and Background

Collaborative improvisation engages experts in physical expression, as well as typical users, in brief, everyday interactions with prototype technologies. We improvise prompts to interact, then examine and discuss their responses, to reveal how people perceive these devices, think and feel about interacting with them, and behave towards them. Through these sessions, we learn about the needs of users and cultures whose backgrounds may vary considerably from our own, along dimensions that include remarkably different physical abilities or limitations, notions of rational behavior or use, or expectations of the technology's sensibility, usability or longevity (Dourish 2012).

Even considering a diverse population of potential users, to each individual, home and work are familiar (and often comfortable) environments, and it can be difficult to question, or evaluate, the role that some novel technology—such as a robotic appliance or furnishing—could take. As Bell et al. (2005) highlight, one challenge in the study of everyday settings is the difficulty of asking questions about what seems obvious [they cite Norman's (1988) discussion of the affordances of glass *"for seeing through and for breaking"*], how this naïve questioning *defamiliarizes the familiar*, and thus supports our efforts to evaluate its significance.

Our approach to conducting improvisation sessions builds upon Gerber (2007), which, in turn, is based upon Johnstone's (1989) exercises to help stage actors perceive and project roles and relationships, respond with spontaneity, and develop skills in narrative storytelling. Gerber maps these concepts over to design activities, and provides examples of how they can be applied to promote collaboration, learning through failure, and skill in presentation—for example, through a group sketching activity where individuals alternatively add single marks. We then implement these concepts through group storyboarding of potential user scenarios, physical and video prototyping, design improvisation sessions, and exploratory lab or field experiments (Sirkin and Ju 2014).

From a methodological perspective, our approach to field and lab studies is similar to the breaching experiments espoused by Garfinkel (1967) as a way to understand the *"seen but unnoticed, expected, background features of everyday scenes,"* and used by Weiss et al. (2008). Weiss and her colleagues deployed an ACE robot in a public shopping area, and approached unsuspecting people with tourist guide information. They then subsequently interviewed people about the interaction to gauge how accepting people were to this discrepant event. This type of event tells us something different from longitudinal deployments of robots in shared settings [a nice survey of such research is found in Leite et al. (2013); a strong case study in Kidd and Breazeal (2008)]: it tells us how people who have incidental and short-term interactions interpret and respond to robot overtures. This class of research has relevance not only to situations where robots are novel, or where robot interactions are passing (Lee et al. 2009), but also to situations where people maybe encountering robots in emergency or disaster situation (Bethel and Murphy 2010), when the nuances of first impressions may be critically important.

Most of our improvisation sessions and experiments employ Wizard of Oz techniques, in which participants are given the impression that they are interacting with fully functional systems, while their interactions are actually mediated by a human operator. Depending on whether activities are focused on ideating or testing, participants may be aware of the human wizard or not. The approach allows both experimenter and participant more freedom of expression, or more systematic constraints, than would be possible with a fully realized system (Dahlbäck et al. 1993). From a design perspective, Wizard of Oz permits the rapid deployment of prototype technologies into naturalistic settings early in the design process, to inform the context of use, to mine the real world for naturalistic social interactions that the device will need to generate and respond to, and to understand critical technical limitations inherent to the application (Riek 2012).

3 Case Studies in Design

Our case studies center on the design of interactions between people and *expressive everyday objects*: specifically, how non-anthropomorphic robots can, and should, interact with people. So what are robotic everyday objects? Our projects include an ottoman that offers to support a seated person's feet, a dresser drawers unit that opens to reveal the right tool just when someone needs it, and a trash barrel that roams around a dining area and collects trash from diners completing their meals. We study such objects *in motion* because interactivity implies sociability, and the mindful design of such near-future autonomous technologies can help avoid social miscues, mismatched conceptual models, and unmet expectations. Over the course of these projects, our goals have been (a) to understand how people respond to novel, agentic devices during their everyday activities, as well as (b) to develop a methodology for creating interactions that read naturally to the people involved. This type of design research takes place within the broader context of embedded computing, smart devices and home automation.

The following sections describe how we designed these three prototypes, and used improvisation to evaluate how best to interact with people, as well as the range of responses these devices received in use. We seek to understand both *normative behaviors*, which are common and expected, and *individual responses*, which are unique and often more idiosyncratic.

Fig. 1 A participant in an improvisation session accepts the ottoman's offer to support his feet

3.1 Mechanical Ottoman

3.1.1 Introduction

The mechanical ottoman is a household robot that approaches a seated person from across a room, offers to support his or her feet, and after doing so for several minutes, requests to take leave of the ongoing interaction. Since the main joint activity in human-ottoman interaction is fairly static, we were primarily interested in the question of how engagement and disengagement occur. We therefore developed the ottoman, and a natural interaction scenario (see Fig. 1), specifically to study how a non-anthropomorphic robot can initiate, participate in, and then disengage from, a joint activity with a human partner (Sirkin et al. 2015).

3.1.2 Prototype Systems

Although we are designing the intended functions and workings of autonomous systems, our approach towards development is to produce quick, inexpensive prototypes that enable us to explore possible forms and functions. We therefore built a low-resolution functional prototype using an inexpensive store-bought ottoman, which we set atop casters, and steered around the floor by hand using 2-m long wooden dowels that we attached to its bottom with gaffer's tape.

We subsequently built a robotic, teleoperated ottoman, modeled on the earlier prototype. We set the ottoman atop a modified Willow Garage Turtlebot, which is based, in turn, on the iRobot Create robotic base. We attached a servo to the topmost portion of the internal frame to raise and lower the ottoman 2.5 cm (vertically), making sure that we reinforced the internal assembly to support the weight of a person sitting on it. A concealed researcher then remotely controlled robot's path and speed across the floor, and its vertical motion.

3.1.3 Designed Behaviors

Through early prototyping and improvisation sessions (described next) we decided that the robot would (a) initiate and conclude each movement with a vertical lift or drop motion, (b) start rolling at a speed of about 1 m/s, (c) approach a seated person following a curving path, always within his or her immediate field of view, (d) pause movement at a distance of 1 m for about 10 s, and (e) resume its path toward the person, slowing to about ½ m/s as it drew near [proxemic movement patterns were informed by Hüttenrauch et al. (2006) and Michalowski et al. (2006)].

If the person did not immediately raise his or her feet, the ottoman would begin a sequence of three increasingly assertive actions. First was a brief lift and drop, next was a quick rotational wiggle around its center, and last was a gentle nudge, or bump, up against the person's legs or feet, alternatively leading from the left or right side. Once someone's feet were actively being supported, the ottoman would then follow a similar sequence to bid to leave.

3.1.4 Improvisation Sessions

There were two distinct phases of design improvisation for the ottoman: an initial, developmental phase using the hand-puppeted prototype, and a subsequent experimental phase using the remotely-controlled robotic prototype.

3.1.4.1 Phase 1: Developing Behaviors

We held three design sessions, each lasting about 2 h, with domain experts in physical movement and interaction, including (a) a dance choreography instructor, (b) an improvisational theater performer and theater director, and (c) a stage actor. We placed these participants in various individual seating arrangements, and *puppeted* the prototype (see Fig. 2), exploring ways that someone could beckon or dismiss the ottoman, what personalities different speeds, gestures or angles of approach and departure projected, and appropriate social distances. We engaged participants in role play, using prompts such as *"shoo the ottoman away as many ways as you can,"* and encouraged them to respond gesturally and speak their reactions aloud.

These sessions often pointed out how social and cultural the interpretations of actions were. At one point, after the robot quickly approached to about a meter away, followed by a pause, then a gentle move closer, the dance instructor declared *"Ah, now we're in India,"* evoking the tradition of exemplary service being viewed as an art form in that culture. She elaborated by contrasting India, or England, to other cultures, such as the United States, which have historical frames of service being viewed as closer to servitude. The improvisational theater performer echoed this sentiment when the ottoman withdrew to a ready position in front and to his

Fig. 2 During an improvisation session, the researcher on the *right* *puppets* the prototype using two wooden dowels

side, stating *"That feels like butlery,"* noting how this action showed an intent to be useful, and allowed it to be recalled to service quickly. The stage actor treated the ottoman more like an animate object, which he could wave over if needed, sneer at if he wanted to cross his legs, or kick gently aside if it encroached too closely into his personal space. Each of these perspectives placed the seated person at the top of the social order, with the ottoman alternatively treated as provider of expert service, obedience or pure functionality.

Several participants felt that a brisk, vertical lift movement of several centimeters (which we called *"stand up"*) suggested attention and a readiness to move, and that the corresponding drop (or *"sit down"*) movement suggested stability, and a likelihood of staying put after completing some action.

3.1.4.2 Phase 2: Exploring and Interpreting

Next, we engaged 20 participants in a lab study, which recreated a natural interaction setting, and explored how they interpreted and responded to the robotic prototype's designed behaviors. We asked participants to sit in a lounge chair and watch a video. After a few minutes, the ottoman approached from a position either nearly in front of them, or off to their side (see Fig. 3). The operator improvised interactions with them, encouraging them to engage, without following a script.

3.1.4.3 Approaching

Almost all participants recognized the ottoman as a robotic footstool, or as one participant described, *"a weirdly sentient footstool."* One or two who were confused at first quickly came to understand its role and intent: *"At some point I thought it could be a small chair or something, but I pretty much got it when it was approaching my feet."* This suggests that not only the robot's form, but its movement, informs how people interpret its role.

Most people accepted the offer to rest their feet on the ottoman, many of them lifting their legs right away as it rolled up toward them (see Fig. 4). About half even

Fig. 3 A typical path taken to approach a seated person. *Squares* represent the ottoman's position, one second apart

Fig. 4 From *left* to *right*: one participant anticipates the arriving ottoman by elevating his feet; another pats it to assuage its offer to support her feet; one rests his feet on it while watching a video; and another avoids interacting with nervous laughter

held their legs suspended in the air for several seconds, so the ottoman could settle down just beneath them. Those who did not rest their feet knew that it was an ottoman, and that it was offering to support their feet, and described their reluctance as perceiving the ottoman to be alive: *"It's a moving thing that I almost perceive as living. I didn't want to denigrate it by using it as a footstool."* Another said, *"Because it felt like it was alive I didn't want to put my feet on it."* and *"I feel like it communicated with me well, but I would feel uncomfortable doing what it was asking."* One other participant rested her feet at first, but removed them after about 30 s, saying later, *"It seemed like it wanted me to put my feet on it, but I didn't want to constrain it too much. I didn't want to imprison it here."*

3.1.4.4 Taking Leave

Everyone who had rested their feet recognized the quick lift-and-drop motion as a request to take leave of the interaction, although some did not notice it until the second or third time: *"It, like, sat up when it was ready to go."* and *"When it signaled to leave, it rose and fell, to let me know it was time to go. I wasn't ready for* *it to leave because I was comfortable. I'm glad it let me know instead of taking off."* Responses to the ottoman's request to withdraw ranged from accepting, *"I think it wanted to go, so I set it free."* to disappointed, *"It made its own decision to leave, like I don't want to be your footrest anymore."* to mildly annoyed, *"I was a little offput when it decided that it wanted to leave. If it was doing that all of the time, then I'm not sure how good of a footrest it would be. I expect a footrest to be there."*

Participants had several rationales to explain its early departure, which ranged from attending to other routine tasks to taking care of someone else: *"I thought it probably had something to do, to go do some errands." "Maybe someone else needed a footrest, or it needed to charge itself."* and *"There was probably somebody more important in the room, so it was going to meet that guy."*

3.2 Emotive Drawers

3.2.1 Introduction

The emotive robotic drawers are designed to help us explore how a robot can participate in an iterative, turn-taking activity with a human collaborator, using its movements alone. The shared activity can be any task where the human needs repeated access to objects stored within the robot's drawers, and where it can open and close in ways that encourage and support task completion. The example that we explored most closely is a cube assembly puzzle (see Fig. 5), which requires six different fasteners, and where the drawers contain all of the tools required to build the cube Mok et al. (2014, 2015).

Fig. 5 Emotive drawers performing an animation gesture while a human collaborator assembles a cube puzzle

3.2.2 Prototype Systems

We fabricated the prototype from a standard IKEA MICKE 4-drawer unit. To allow the drawers to perform consistent and repeatable motions, we retrofitted the top three drawers with motors and a rack-and-pinion system. Spring-loaded rotary encoders mounted against the unit's frame allowed us to track each drawer's position. The bottom drawer was taped off from access during the study and contained parts and hardware that drove the system. Actuation was controlled by an Arduino microcontroller communicating with a local client program over USB cable.

The local client program provided a Wizard-of-Oz style remote control for the drawers. A researcher in an adjacent room operated the drawers via keyboard hotkey with 15 buttons, each of which was bound to a pre-programmed sequence of movements, or *"animations,"* for the drawers to execute. The researcher had a one-way video feed of both the drawers and participants, which was used to observe participants' actions and improvise appropriate responses.

3.2.3 Designed Behaviors

Animations were modeled to be either non-expressive or expressive. The non-expressive case used only the most basic, functional movements, and included *simple open* (Fig. 6a), where each drawer opened at a constant speed, and *simple close* (Fig. 6b), where all of the drawers closed at the same constant speed and locked after closing.

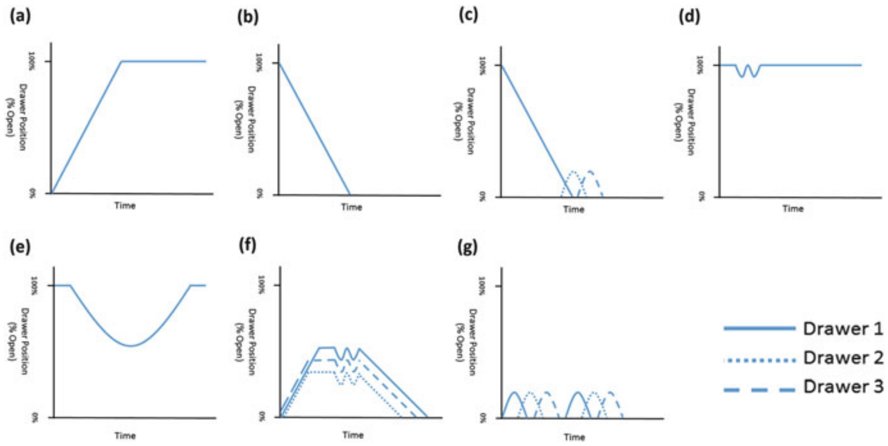


Fig. 6 Drawer animations include (**a**) simple open, (**b**) simple close, (**c**) flair close, (**d**) wiggle, (**e**) beckon, (**f**) chuckle and (**g**) happy

The expressive case added flair to the closing animation, and more communicative gestures. For *close flair* (Fig. 6c), after one drawer closed, the other two would open and close slightly in sequence, mimicking a ripple effect. Other animations were designed to suggest to the participant that he or she return or remove an item from an open drawer. For the former, the drawers would *wiggle* (Fig. 6d), where an open drawer moved in and out slightly, but quickly, two times. For the latter, the drawers would *beckon* (Fig. 6e), where an open drawer closed halfway, and then reopened at half speed. Two final animations suggested positive sentiment. For *chuckle* (Fig. 6f), all of the drawers opened to random positions, wiggled twice and then closed. For *happy* (Fig. 6g), all of the drawers mimicked a ripple effect, which traveled down the drawers unit twice.

3.2.4 Improvisation Sessions

Twenty people participated in a lab study, which simulated a mechanical workspace within an office environment, to interact with the drawers and complete the cube puzzle. For each session, the robot was either non-expressive or expressive. Additionally, it took on one of two levels of assertiveness: *proactive*, where the drawers led the activity by initiating actions, or *reactive*, where the drawers waited for gestures by the participant before responding. We gave participants very little guidance about how to engage the drawers, hoping that they would negotiate an approach that made sense to them. The operator could then freely improvise the robot's behavior, using the animations available for that session's persona.

3.2.4.1 Proactive Action

When the drawers were proactive, participants felt that the robot was not treating them as equals. To one, the robot appeared very much to be *"Like a boss."* figure that relegated people to a lower social standing. Another indicated that *"I was the builder, the drawers should not command me to do things. I will do it when I am ready."* Yet another felt that *"It was distracting when I was trying to understand what's going on. I knew it was trying to get me to get the tool for the next step, but I didn't know what to do yet. So I just ignored it until I was ready."* Conversely, when the drawers were expressive, they did not create this feeling of inequality. One participant noted, *"It was like a fiery little Scottish Terrier trying to pull me its way."* Several others similarly noted that it was *"Like a pet."* So, by incorporating an expressive nature into its actions, the robot can still help lead interactions, while making participants feel more like equals, and assuaging the frustration and discomfort that they might otherwise feel.

3.2.4.2 Expressive Movement

The expressive movement displays were also effective at making an impression on participants, regardless of their immediate focus of attention. For example, participants' attention was often drawn somewhere other than the drawers—say, toward the cube puzzle task—while an expressive drawer performed an animation. Yet they later recall having seen the animation, with a large percentage realizing that *"It was trying to congratulate me."* Regarding the *chuckle* animation, one participant noticed, *"It was warning me of an error."* So, even without looking directly at the drawers, participants still experienced their movements. This reaffirms Hoffman and Ju's (2014) finding that people are sensitive to movement, and that small, well-designed motions can be used to communicate.

3.3 Roving Trash Barrel

3.3.1 Introduction

We built and tested a roving trash barrel robot to better understand the implicit social protocols, cues and signals of public interactions between people and robots, and to produce movements and actions that people can understand. In eight lunchtime sessions, spread across two heavily populated campus dining destinations, we piloted the robot in Wizard-of-Oz fashion through crowded public areas, initiating and responding to requests for impromptu interactions around collecting people's trash (see Fig. 7).

Fig. 7 A lunchtime diner discards his trash into the roving trash barrel after it approached and gestured to draw his attention

3.3.2 Prototype Systems

The trash barrel robot was designed to resemble those commonly found around the university's campus. Its body is a standard 32 gal BRUTE gray trash barrel from Rubbermaid's line of commercial products (see Fig. 8). The barrel is mounted atop a robotic base, powered by an iRobot Create, and augmented with a laptop computer (concealed within the trash barrel), two web cameras (hidden within the barrel's grab handles) and a microphone (mounted under the barrel's topmost lip).

A researcher remotely controls the robot (Fong and Thorpe 2001), issuing commands over a web interface to a remote server, which relays the control commands over WiFi to the laptop, and then by USB to the robotic base (see Fig. 9). The interface shows the operator two video streams from the cameras in the robot's handles. The choice of teleoperation over autonomous control provides flexibility for real-time improvisation, and permits responsive behaviors to unanticipated events.

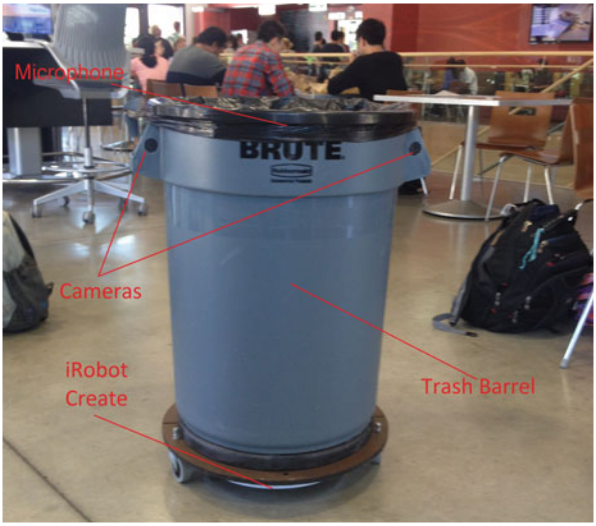


Fig. 8 A standard Rubbermaid Brute trash barrel is mounted atop an iRobot Create platform. A laptop computer, hidden within the trash barrel, handles video and control commands

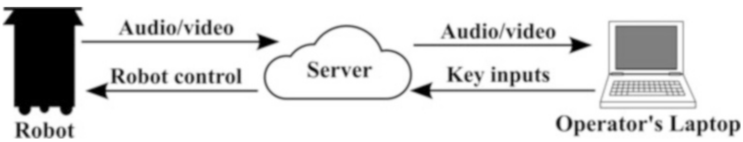


Fig. 9 The roving trash barrel robot's control system. Cameras on the robot send video to a remote operator's interface on a laptop, which sends motion commands back to the robot

3.3.3 Designed Behaviors

Owing to its iRobot Create drivetrain, the trash barrel could move forward and backward, straight or along an arc, or rotate in place: it was incapable of lateral motion without first rotating. This limitation, combined with the use of cameras to view and interact with people, created an implicit front for the otherwise cylindrical trash barrel, and meant that the robot *"faced"* a person or group as it approach them.

Aside from speed and direction control, the robot had three pre-programmed behaviors: *wiggling*, where it quickly rotated left and right, *nudging*, where it abruptly moved front and back, and *beeping*, where it played a neutral two tone beep. At high speed, the *wiggling* and *nudging* movements appeared more like *shivering* and *nodding*, respectively. The combination of basic drive and programmed behaviors allowed the operator to maneuver the trash barrel, and signal intent, improving the legibility and predictability of its actions (Takayama et al. 2011; Dragan et al. 2013).

3.3.4 Improvisation Sessions

We conducted eight sessions, during lunch at two busy dining locations, in which we explored different ways to approach, gesture, disengage and acknowledge people. Because our goal was to elicit naturalistic behavior, we did not pre-warn potential participants that they could be engaged in an improvisational interaction, but we did interview those who actively engaged with the robot afterward (155 in total). Due to the crowd and noise, people were rarely aware of the trash barrel until it approached within an arm's length distance. It could therefore wander throughout the area, interacting opportunistically with nearby people along its path, without their anticipating its arrival.

There were two phases of improvisation. In the first, exploratory phase, the trash barrel *wiggled*, *nudged*, *bumped* and *beeped* to attract attention. It could even interrupt conversations and make excessive noise by dragging along empty chairs. In the second, goal-oriented phase, the trash barrel followed a loose script to encourage people to discard their trash. Scripts called for the robot to either (a) visit every table along a set path and enact a simple stop-and-move-on behavior at each one, (b) initiate and respond to requests to engage people by responsive movement, or (c) purposely appear to struggle, by bumping into chairs or uneven pavement, with the intent to elicit empathy.

3.3.4.1 Interrupting Activity

People who were alone were less likely to engage socially with the robot, but if several people in a group noticed it, they would engage extensively. Overall, people did not readily interrupt ongoing activities, whether they were watching a video alone or socializing with friends. This correlates with results by Hüttenrauch and Hüttenrauch and Severinson-Eklundh (2003), who found that people do not attend to robots when they are already engaged in their own activities.

3.3.4.2 Overt Non-interaction

Most people appreciated the robot's presence and offer, particularly when they were in small social groups, where the person closest to the robot could check the responses of the others before deciding that it was okay to engage. The most common way for people to signal their unwillingness to interact was to ignore the robot's presence (Fischer et al. 2014), although some people overtly avoided interacting by averting their gaze or turning their faces away, and not responding to the robot's actions as it drew near. One woman kept a hand by her face, as if to hold her hair back from her eyes, but always at an angle that kept her from visually connecting with the robot. Others avoided looking toward the robot as it approached, but turned around and recorded video as it departed to interact with others. Still others performed short, curt interactions, and then pointedly turned away to indicate that the engagement was over. More observations from this study have been written up in Fischer et al. 2015.

3.3.4.3 Ascribing Desires

People often waved to attract, or shooed to dismiss, the robot. The most common gesture to call it over was to wave trash in its line of sight. This pattern is consistent with the material signals that people use when coordinating joint action, as pointed out by Clark (2005). But we also observed actions that went beyond just signaling for coordination. People appeared to ascribe desires and motivations to the trash barrel, that it intrinsically *desires* trash rather than just performs a fixed collection routine. One demonstration of this belief involved a family with a 5-year-old son. Over 30 min, as the robot interacted with others, the child re-approached the trash barrel on four occasions, waving trash directly in front of its cameras before slowly pulling the trash back, in an attempt to coax the trash barrel to follow him, as if the boy's mental model of the robot included its intent to consume trash, despite his never having been informed as such. Adults also seemed to hold this mental model. Several described throwing away trash as *"feeding"* the trash barrel, and expressed a desire for it to acknowledge, or thank, them for having given trash to it.

3.3.4.4 Empathy and Altruism

Both accidentally and intentionally, the trash barrel made a series of mistakes and exhibited struggling behavior—where it repeatedly tried to reach an objective, but was unable to overcome obstacles, such as furniture or pavement edges. Some people felt that the robot was *"Not very smart."* and exhibited *"Erratic behavior."*

But others found struggling to be endearing, characterizing the robot as being like *"A puppy or a toddler."* One said, *"When it ran into the garbage cans, I thought, 'silly robot!' It was adorable."* This attitude appeared to encourage them to move obstacles, such as chairs, out of the robot's way. When asked later why she had done so, one person replied, *"I don't know, it felt like it was a team effort."* Another person *"Noticed that the robot made people laugh and smile, and when it got stuck, I wanted to help it because I thought, 'the show must go on.'"*

4 Next Steps

The subject of our *embodied design improvisation* approach has been to build interactive robots that co-inhabit spaces with people. As a next step in extending this method, and in order to design gestures and movements that are more meaningful and socially appropriate, we propose to utilize the *"wisdom of crowds"* to help teleoperate robots. We are interested in better understanding when the crowds are indeed wise, and when they need assistance. We would like to know how best to aggregate crowd input, to handle transfers of control, and to give remote workers situational awareness.

We expect that experimentation with this *crowdpowered embodied interaction* approach will help us to elicit context and culture-sensitive rules of interaction that are key to designing new products; that these techniques will lead to insights that are more easily codified into machine action than those generated from live, in-person improvisation; and that end users will respond more positively to ideas generated from more people than to those developed without the benefit of as much collaboration.

References

Bell G, Blythe M, Sengers P (2005) Making by making strange: defamiliarization and the design of domestic technologies. ACM Trans Comput Hum Interact 12(2):149–173

Bethel C, Murphy R (2010) Emotive non-anthropomorphic robots perceived as more calming, friendly, and attentive for victim management. In: AAAI fall symposium: dialog with robots

Clark H (2005) Coordinating with each other in a material world. Discourse Stud 7(4–5):507–525

Dahlbäck N, Jönsson A, Ahrenberg L (1993) Wizard of Oz studies—why and how. Knowl Based Syst 6(4):258–266

Dourish P (2012) Designing for zombies. Unpublished manuscript available at http://www.douri.sh/publications/2012/zombies.pdf

Dragan A, Lee K, Srinivasa S (2013) Legibility and predictability of robot motion. In: Proceedings of the ACM/IEEE international conference on Human-Robot Interaction (HRI 2013), ACM, pp 301–308

Fischer K, Soto B, Pantofaru C, Takayama L (2014) Initiating interactions in order to get help: effects of social framing on people's responses to robots' requests for assistance. In:

Proceedings of the IEEE international symposium on robot and human interactive communications (RO-MAN 2014), IEEE, pp 999–1005
Fischer K, Yang S, Mok B, Maheshwari R, Sirkin D, Ju W (2015) Initiating interactions and negotiating approach: a robotic trash can in the field. In: Proceedings of the AAAI 2015 spring symposium on turn-taking and coordination in human-machine interaction, Palo Alto, CA
Fong T, Thorpe C (2001) Vehicle teleoperation interfaces. Auton Robots 11(1):9–18
Garfinkel H (1967) Studies in ethnomethodology. Prentice-Hall, Englewood Cliffs, NJ
Gerber E (2007) Improvisation principles and techniques for design. In: Proceedings of the SIGCHI conference on human factors in computing systems (CHI 2007), ACM, pp 1069–1072
Hoffman G, Ju W (2014) Designing robots with movement in mind. J Hum Robot Interact 3 (1):89–122
Hüttenrauch H, Severinson-Eklundh K (2003) To help or not to help a service robot. In: Proceedings of the IEEE international symposium on robot and human interactive communications (RO-MAN 2003), IEEE, pp 379–384
Hüttenrauch H, Severinson-Eklundh K, Green A, Topp E (2006) Investigating spatial relationships in human-robot interaction. In: Proceedings of the IEEE/RSJ international conference on Intelligent Robots and Systems (IROS 2006), IEEE, pp 5052–5059
Johnstone K (1989) Impro: improvisation and the theatre. Methuen, London
Kidd C, Breazeal C (2008) Robots at home: understanding long-term human-robot interaction. In: Proceedings of the IEEE/RSJ international conference on Intelligent Robots and Systems (IROS 2008), IEEE, pp 3230–3235
Laurel B (2003) Design improvisation: ethnography meets theatre. In: Laurel B (ed) Design research: methods and perspectives. MIT Press, Cambridge, MA, pp 49–54
Lee M, Forlizzi J, Rybski P, Crabbe F, Chung W, Finkle J, Glaser K, Kiesler S (2009) The snackbot: documenting the design of a robot for long-term human-robot interaction. In: Proceedings of the ACM/IEEE international conference on Human-Robot Interaction (HRI 2009), ACM, pp 7–14
Leite I, Martinho C, Paiva A (2013) Social robots for long-term interaction: a survey. Int J Soc Robot 5(2):291–308
Michalowski M, Sabanovic S, Simmons R (2006) A spatial model of engagement for a social robot. In: Proceedings of the international workshop on advanced motion control, IEEE, pp 762–767
Mok B, Yang S, Sirkin D, Ju W (2014) Empathy: interactions with emotive robotic drawers. In: Proceedings of the ACM/IEEE international conference on Human-Robot Interaction (HRI 2014), ACM
Mok B, Yang S, Sirkin D, Ju W (2015) Every tool in its place: Interaction and collaboration with robotic drawers. In: Proceedings of the AAAI 2015 spring symposium on turn-taking and coordination in human-machine interaction, Palo Alto, CA
Norman D (1988) The design of everyday things. MIT Press, London
Riek L (2012) Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. J Hum Robot Interact 1(1):119–136
Sirkin D, Ju W (2014) Using embodied design improvisation as a design research tool. In: Proceedings of the international conference on Human Behavior in Design (HBiD 2014), Ascona, Switzerland
Sirkin D, Mok B, Yang S, Ju W (2015) Mechanical ottoman: how robotic furniture offers and withdraws support. In: Proceedings of the 10th ACM/IEEE international conference on Human-Robot Interaction (HRI 2015), ACM
Takayama L, Dooley D, Ju W (2011) Expressing thought: improving robot readability with animation principles. In: Proceedings of the ACM/IEEE international conference on Human-Robot Interaction (HRI 2014), ACM, pp 69–76
Weiss A, Bernhaupt R, Tscheligi M, Wollherr D, Kuhnlenz K, Buss M (2008) A methodological variation for acceptance evaluation of human-robot interaction in public places. In: Proceedings of the IEEE international symposium on robot and human interactive communications (RO-MAN 2008), IEEE, pp 713–718

Part II
Creativity and Creative Confidence

Designing a Creativity Assessment Tool for the Twenty-First Century: Preliminary Results and Insights from Developing a Design-Thinking Based Assessment of Creative Capacity

Grace Hawthorne, Manish Saggar, Eve-Marie Quintin, Nick Bott, Eliza Keinitz, Ning Liu, Yin-Hsuan Chien, Daniel Hong, Adam Royalty, and Allan L. Reiss

Abstract In order to assess a person's creative capacity in real-world situations, we propose a novel Design Thinking Creativity Test (DTCT). The DTCT is based on the design thinking principles and can serve as an assessment that reflects problem-solving needs of the twenty-first century. In particular, the DTCT emphasizes assessment of case-based skills to directly measure an individual's application of creativity during an innovation event. In order to create a robust and standardized creativity assessment, we are currently examining (1) the reliability of DTCT in capturing applied creativity in a simulated real-world setting; (2) the validity of DTCT by relating DTCT scores with other standardized assessments of creativity; and (3) the correlations between neuroimaging data and scores achieved in the DTCT task. In this chapter, we provide a thorough background on already existing tools to assess creative capacity and how our approach advances the current state of the art. We also share challenges faced in collecting and analyzing DTCT data, along with proposed solutions. Lastly, we provide our hypotheses and preliminary insights regarding DTCT's ability to capture applied creative capacity.

G. Hawthorne (✉) • A. Royalty
Hasso Plattner Institute of Design (d.school), Building 550, 416 Escondido Mall, Stanford, CA 94305-3086, USA
e-mail: grace@dschool.stanford.edu; aroyalty@stanford.edu

M. Saggar • E.-M. Quintin • N. Bott • E. Keinitz • N. Liu • Y.-H. Chien • D. Hong • A.L. Reiss
Center for Interdisciplinary Brain Sciences Research, Stanford University School of Medicine, 401 Quarry Road, Stanford, CA 94305-5795, USA
e-mail: saggar@stanford.edu; quintin@stanford.edu; nbott@stanford.edu; ekienitz@stanford.edu; ningl@stanford.edu; yhchien1@stanford.edu; weichenh@stanford.edu; reiss@stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_8

1 Introduction

Outside of popular, but dated, psychometric instruments like the Torrance Test of Creative Thinking (TTCT) (Torrance 1990) and Alternative Uses Test (Mendelsohn 1976), a widely applicable creativity measurement that takes into account new information and concepts from recent behavioral- and neuroscience research has yet to emerge from recent creativity studies. While the popular TTCT has picture- and verbal-based exercises to assess an individual's possession of mental creativity characteristics and divergent thinking abilities, it does not address the contemporary construct of creativity and innovation that includes associative abilities and an individual's ability to overtly apply their creativity to real world scenarios.

In order to better assess a person's creative capacity in real-world situations, we developed a novel Design Thinking Creativity Test (DTCT) as a next generation creativity assessment that reflects problem-solving needs of the twenty-first century. In particular, the DTCT emphasizes assessment of case-based skills so as to directly measure an individual's application of creativity during an innovation event.

The Hasso Plattner Institute of Design at Stanford University (Stanford d.school) offers classes specifically aimed at enhancing creative capacity through design thinking skill building. As is custom in the academic tradition, instructors must come up with ways to evaluate their student's progression. Often, students are asked to produce deliverables that are then judged by the instructors and classmates. Although this method has academic value, its purely qualitative nature does not allow a formal assessment or measurement on whether the student's creative capacity has been enhanced. In addition, merely possessing creative capacity is differentiated from having the ability to activate and deploy creative capacity. Thus, it is important to examine whether an individual's creative capacity can be objectively measured in real-world settings and enhanced.

Thus, our primary question is: Can an individual's creative capacity be objectively measured in real-world settings and enhanced, and if so, what essential characteristics should be part of the measurement and how does it align with changes in brain functioning?

The development of a new creativity assessment prompted other questions that will be addressed in this chapter. How reliable is a case-base format in capturing applied creativity in a simulated real-world setting? How do results from test subjects' performance on a creativity assessment correlate with functional neuroimaging data both pre- and post-creativity training? How do the results from the DTCT assessment map to the results from the TTCT assessment? Can a case-based creativity assessment measure the application of creativity skills?

The creation of a statistically robust, well-standardized test relies on the acquisition of large amounts of data during large-scale studies in which representative groups of individuals take the test under standardized conditions (Zucker 2003). In this project, we will align the DTCT with creativity training goals from the design thinking methodology as our initial attempt to establish face validity. By doing so, we hope to draw connections between creativity training, its effect on an individual's creative capacity, and the potential impact of increased capacity from training. By attaining this goal, we (and others) will be able better evaluate the effects of training/teaching methodologies by utilizing a direct, twenty-first century assessment instrument.

We have already laid preliminary groundwork for a next generation creativity assessment through research efforts from a collaboration between Stanford's Hasso Plattner Institute of Design and Stanford School of Medicine's Center for Interdisciplinary Brain Sciences Research (CIBSR), which were focused on: (a) how creativity is manifested in the brain; (b) if training can enhance creative capacity; (c) what are the associated neural correlates of such enhancement; and (d) whether repeated training is required to maintain the enhancement of creative capacity. This was performed through a rigorous longitudinal study that included two 5-week long interventions/trainings sessions and fMRI brain scans, neurocognitive, and behavioral assessments at three time points (Bott et al. 2014; Hawthorne et al. 2013; Kienitz et al. 2014; Saggar et al. 2014).

2 Assessing Creative Capacity

In this section, we briefly provide a brief definition of creativity as well as thorough background on the previous tools used to assess creative capacity in healthy individuals. These tools are broadly divided into following five categories: (a) convergent thinking based tools; (b) divergent thinking based tools; (c) artistic assessments; (d) self-assessments; and (e) improvisation based tools.

2.1 Creativity in the Twenty-First Century Defined

For the purpose of this study, our definition of creativity is guided by the philosophy of the Stanford d.school while insuring translation to the field of cognitive neuroscience. We define creativity as "*a state of being and adaptation of personal skill sets that enables an individual to synthesize novel connections and express meaningful outcomes*" (Hawthorne et al. 2013). This definition captures the intersection of three different axes. To determine how creative a person, deliverable or process is, these components can be rated along three continuums from—(a) existing to new/novel, (b) linear to synthesizing, and (c) no value/meaning to meaningful. We propose a visual illustration of these continuums with three axes (Fig. 1a). A deliverable or process with high novelty, meaning, and synthesis is considered highly creative and so is the person responsible for this deliverable or process. This person, the deliverable, or the process falls within the upper right and back zone of

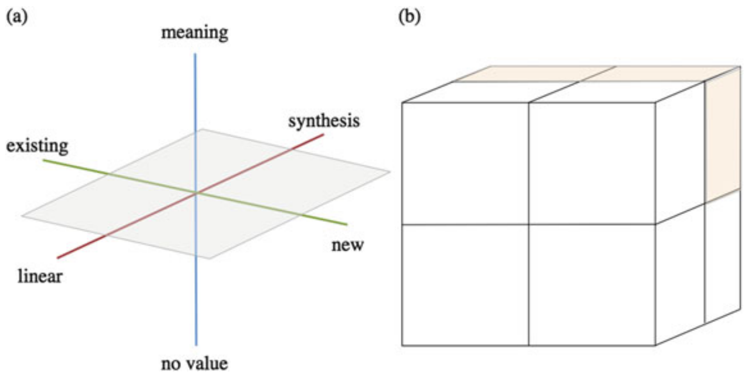


Fig. 1 Visual illustration of our working definition of creativity

the three-dimensional space created by these three axes (the zone in orange in Fig. 1b).

This definition of creativity focuses attention to the person and their skills as opposed to process and outcomes as more traditionally defined. The intention of this focus is to better align the skill of creativity to indications that go beyond the possession of creativity into the ability to exercise/apply it.

2.2 Convergent Thinking Tests

In the extant literature, convergent thinking is defined as the ability to provide the single correct solution to a given problem (Runco 2004). This ability to come up with the right solution, given the constraints of a problem, is thought to require (a) taking a novel approach to the problem; (b) seeing the problem from a different perspective; or (c) making a unique association between the given/associated ideas. The solutions are usually scored in a binary fashion, as either correct or incorrect.

To test the convergent thinking capacity, researchers usually give participants a set of insight (or "aha") problems (Dow and Mayer 2004). The insight problems are different from the broader class of well-structured problems, such that the solution to an insight problem specifically requires a mental set shift or reconceptualization of the problem and given constraints (Goel 2014). Once such mental shift occurs, the problem solver can access the solution with ease. For example, consider the famous insight problem—the triangle (Katona 1940). In this problem, the goal is to arrange six sticks of equal length into four equilateral triangles. Until an individual considers using the third-dimension of space, he/she cannot solve this problem. However, once the shift occurs, and he/she realizes the use of three dimensions, instead of two, the solution presents itself with ease. This phenomenological shift in solving the problem is also referred to as having an "aha" moment.

Another method of assessing convergent thinking capacity is by using the Remote Association Test (RAT; Mednick 1962). In the RAT, participants are presented with three words and are asked to come up with a single word that is associated with all the three given words. For example, when given the three words—"cottage, blue, and mouse", the expected and correct solution is "cheese". As in the case of insight problems, the RAT solutions are also scored in a binary fashion, i.e., as right or wrong.

In summary, convergent thinking tests assess participant's ability to converge to a single correct solution by shifting their mental perspectives. Such thinking, although crucial for creativity, cannot fully capture the creative process.

2.3 Divergent Thinking Tests

Since J. P. Guilford's seminal lecture on motivating research on creativity in 1965, most (if not all) studies have focused on assessing creativity in the form of divergent thinking (Dietrich and Kanso 2010). Divergent thinking is defined as the ability to rapidly generate several ideas/solutions for a given problem. In contrast to convergent thinking, divergent thinking problems are open-ended and do not have one exact solution. Due to the open-endedness, solutions for divergent thinking problems are scored on different putative components of creativity. In the extant literature, some of the widely used components are—(a) originality, i.e., statistical infrequency of the response); (b) fluency, i.e., number of responses; (c) flexibility, i.e., number of unique classes of responses; and (d) elaboration, i.e., the amount of detail in the response.

One of the most widely used tests of divergent thinking is the Torrance Test of Creative Thinking (TTCT; Torrance 1990). The TTCT is available in both verbal and figural versions. The TTCT figural is more widely used as it is less influenced by vocabulary and is appropriate at all levels of age (kindergarten through adult). In TTCT-Figural, participants are given an incomplete figure and are asked to complete the figure in a given amount of time. Participants are also told to complete the figure in such a way that it would portray the most "unusual and unique" story. Trained staff members at the Scholastic Testing Services (STS) later rate the final drawings.

Other divergent thinking tests include Wallach and Kogan's assessments of creative capacity (Wallach and Kogan 1965). In this test participants are asked come up with as many possible items as they can in specific amount of time for a given component. For example, if the given component is "wheels", participants are asked to come up with as many items as they can which contain the component wheels. Like the TTCT, Wallach and Kogan's test is also rated on flexibility, fluency, originality, and elaboration sub-scales.

Other examples of divergent thinking tasks include Alternate Uses Test (AUT; Guilford 1950), where participants are asked to come up with as many unusual or alternate uses of a given item in fixed amount of time. For example, participants could be asked to rapidly provide alternate uses for a “pen”.

In summary, divergent thinking tests are designed to assess a participant’s ability to rapidly respond and produce as many unusual or original ideas as possible. It is important to note, however, that the focus in divergent thinking tests is on quantity and not quality/utility of ideas. Thus, divergent thinking is focused on a specific, albeit crucial, idea generation aspect of creativity.

2.4 Artistic Assessments

In addition to testing the creativity process, via convergent and divergent thinking, researchers have also focused on assessing the end product for its creative content. Such assessments are commonly done for artistic products, e.g., paintings, story, poem etc. Two or more field experts typically perform evaluations of such artistic products.

As artistic assessments are limited to professional artists or creative individuals, the Barron-Welsh Art Scale (BWAS) was developed to capture similar assessments in nonprofessional artists (Barron and Welsh 1952). The BWAS is a figure preference test, which includes simple and complex figures that are liked and/or disliked by established professional artists. The participant is shown these figures and is asked to give his/her preference about their liking/dis-liking of the figure. An individual scores higher on BWAS if his/her liking is inline with the artist’s preference (Gough et al. 1979).

Although BWAS scores of creativity have been shown to be associated with other indicators of creativity (Gough et al. 1979), there is strong controversy surrounding BWAS’ validity and the specific dimension of creativity assessed (Ridley 1979).

2.5 Self-Assessments

In addition to assessing creative process and product, researchers have also investigated different aspects of the creative person him/herself using self-assessments. Such self-assessments measure a participant’s creative confidence in generating novel ideas and employing out-of-the-box thinking. Several self-assessment scales were developed over the last 50 years, including Runco’s ideation and behavior scale (Runco et al. 2001), the Creative Attitude Survey (Schaefer and Bridges 1970), the Creative Achievement Questionnaire (Carson et al. 2005), and more recently, the Creative Confidence Scale (Royalty et al. 2014).

Self-assessment scales provide a unique indicator for “personal” aspects of creative thinking. However, such scales by themselves are not enough to measure creative capacity and are generally used in conjunction with other measures of creative thinking.

3 Designing a Design Thinking Based Creativity Assessment

To design the next generation assessment of creativity, we utilized the conceptualization of creativity as prescribed by the Stanford d.school. Further, we narrowed down to specific attributes or characteristics that are critical to creativity and innovation in the twenty-first century. Next, we looked at the landscape of other existing creativity assessments to determine applicability and deficiencies as related to our measurement objective. By determining which cognitive and psychological attributes are measured by existing assessments and whether they addressed our definition of creativity, we saw an unaddressed opportunity to assess creative skill activation versus possession. Existing assessments largely overlook applied or real-world creative acumen that allow participants to demonstrate basic design thinking practices including need finding, creating a point of view, ideating, prototyping, and iterating within constraints or in the face of changing conditions. Thus, an important goal in developing the novel DTCT assessment was to fill these gaps and develop a creativity assessment to measure applied and real-world creative acumen. In this section, we provide details about DTCT's design and its components.

3.1 Case Based Assessment

A case based approach that allows participants to deploy creative skills with intention opens up the potential of assessing creative capacity in a manner that goes beyond the domains of divergent or convergent thinking. By including the creation and building of a real solution from a provided set of cues within a given set of constraints, participants have the opportunity to create a deliverable with high novelty, meaning, and synthesis within an assessment framework. But similar to existing divergent and convergent thinking assessments, in this case based approach, participants also have the opportunity to both conceptualize numerous possible solutions and narrow on one particular outcome of their choosing during the assessment. The case based approach that we have created is a next generation hybrid that builds on pertinent aspects of previous creativity assessments. Our case based approach focuses on assessing applied thinking and adaptability of solutions when faced with new constraints.

Fig. 2 DTCT assessment steps and description

STEP	ACTIVITY
1	Condition evaluation
2	Condition identification
3	Need distillation
4	Point of view ideation
5	Point of view distillation
6	Prototyping
7	Expression/communication with words
8	Expression/communication with pictures
9	Condition change and recast condition
10	Prototyping
11	Point of view distillation
12	Debrief

3.2 DTCT Components

The DTCT was developed around Stanford d.school's design project structure, a fast-paced approach that runs the participant through a full design cycle. Instead of pairing up with another person in a typical design project, participants administered the DTCT run through a course of prompts and activities to create a design solution that is "useful and meaningful" to a hypothetical user.

In the course of 30 min, participants employ characteristics of applied thinking and adaptability for need-finding as well as developing novel solutions. The "case study" assessment begins with a photographic image that consists of people in an environment. Three versions of the DTCT were created for up to three different assessment points; all have the same twelve steps and prompts, and vary in the opening environment and related condition change (Fig. 2).

4 Administration and Data Collection

We prototyped and administered the DTCT as part of another study focused on individual creative capacity building (Hawthorne et al. 2013). Specifically, the DTCT was administered along with several other assessments and MRI scans to 36 participants at three different time points during the 12-week study.

The challenging aspect of administering this assessment is that the activities are open-ended and involve several stages of need finding and brainstorming, which lead to couple steps requiring the development of three dimensional prototypes. To develop 3-D prototypes, all participants were provided with a similar set of limited supplies (e.g., tongue depressors, tin foil, pipe cleaners, post-its, etc.). To capture and later on rate each prototype, photographs of the prototypes were taken after completion of the respective prototyping steps. In our prototype, each study participant did the DTCT on three different occasions over three measurement periods during a 12 weeks period. Because the structure of the assessments were similar, multiple versions could be administered simultaneously.

A large amount of multi-dimensional data already collected from the aforementioned creative capacity study will allow us to refine and improve the DTCT by investigating associations between different domains of the DTCT and other measures of creativity, cognition, personality and brain function.

4.1 Assessment Timing

We collected data related to creativity, cognition, behavior and brain function before (time 1 or T1) and after (time 2 or T2) a creative capacity building intervention based on the Creative Gym class offered at Stanford's d.school (link: http://dschool.stanford.edu/creative-gym-a-design-thinking-skills-studio/). We included a parallel control group receiving a Chinese language and character drawing learning intervention ("Language Gym") as a non-creativity enhancing control condition. Each group "Crossed-over" to the alternate intervention followed by behavioral assessments and brain scans at time 3 (or T3). Therefore, the DTCT was administered to both groups at three different time points (T1, T2, T3) (Fig. 3).

4.2 Assessment Conditions

The assessment was administered by two facilitators in a group setting in an open room. Each participant was assigned a number and given a version of the DTCT activity booklet that they had not previously received. Each step had an assigned time allotment that was monitored by the facilitators, who were available to answer any questions and capture photographs of the 3-D prototypes built by participants. Photos of the built prototypes had to be accurately labeled during the assessment to ensure the images could be properly matched to a respective participant's assessment booklet.

5 Data Analysis

The data collected during each DTCT assessment include photographs of the prototypes developed, as well as in process thoughts and designs scribbled on the paper booklet. After de-identifying and scanning the booklets and photographs, a web-based assessment tool was generated for each participant. Using this

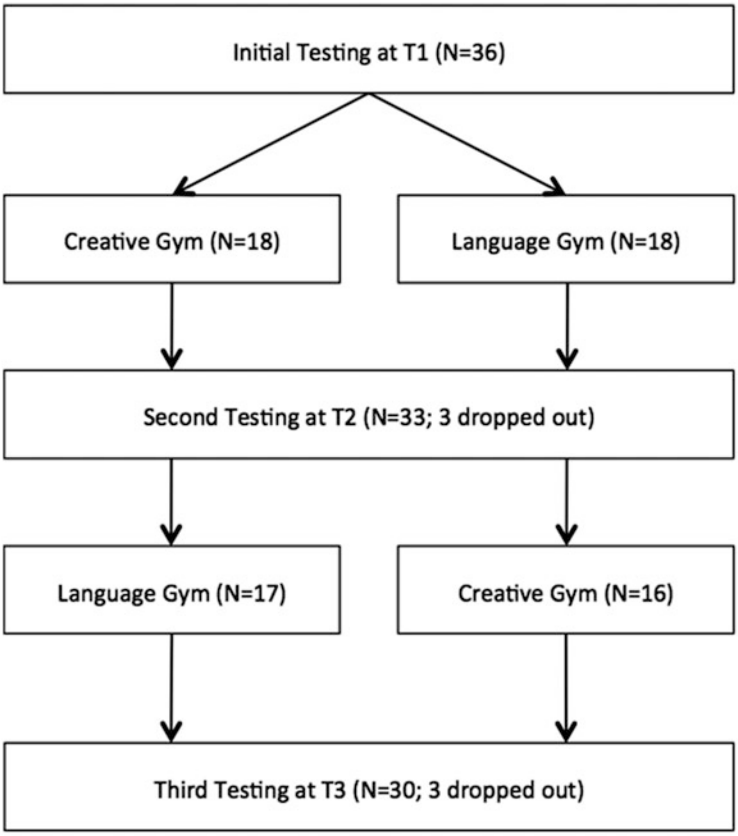


Fig. 3 Visual overview of DTCT administration timing

web-based tool, the raters (at least two) assess and evaluate the work in an efficient, reliable and economic fashion. To validate the novel DTCT assessment, we are currently conducting the following three analyses.

5.1 Convergent Validity

Convergent validity analysis (CVA) will be performed to examine the degree to which the novel DTCT assessment is theoretically related to other standardized and established assessments of creativity (e.g., TTCT). The CVA is usually estimated by correlating scores from two different assessment tools. High CVA scores generally indicate a successful convergence and thereby provide evidence for measuring same/similar underlying concept using the novel and previously established tools. For CVA we plan to use the time 1 data, i.e., data from all 36 participants before intervention.

5.2 Test–Retest Reliability

The reliability of the DTCT over time is an important component of establishing the overall validity of the tool. We will utilize the individual DTCT scores from participants before and after they receive the Language Gym to perform test–retest reliability. We hypothesize that 5 weeks of Language Gym should not significantly affect an individual's creative capacity. Thus, high correlation values between DTCT scores before and after Language Gym are expected.

5.3 Discriminative Validity Analysis

To examine how much value DTCT adds above and beyond the standardized tests of creativity (e.g., TTCT), discriminative validity analysis (DVA) will be performed. In this approach, we hope to show that the DTCT provides additional information about the cognitive skills associated with creativity as well as intervention related outcomes. For DVA, we will use measures of executive functioning, intelligence, personality, as well brain imaging data.

6 Establish a Scoring Guide and Norms

The standardized administration, scoring procedures and norms, and the development and evaluation of the DTCT will begin with already administered tests. Given the objectives and steps included on the test, norm-referenced measures of creative capacity and application will be scored by a host of categories derivative of the TTCT's widely used scoring workbook (Ball and Torrance 1984). The categories will be constructed around key design thinking attributes in the assessment. We will ask the Stanford d.school's most gifted design thinkers to take and score the test to establish the highest possible composite score. We will apply the prototype scoring guide and norm to the already administered tests from the creativity capacity study (108 tests) and to the wider test population that will take the DTCT for this study to refine reliability and validity of the assessment.

7 Looking Ahead

Wider testing of the DTCT prototype will be executed to refine assessment format and improve simplicity and test structure. This will include assessment of the reproducibility of the instrument by administration of alternate forms of the DTCT to subjects within a proscribed period of time. In addition, full analysis of newly administered DTCT test scores to additional subjects will be performed and correlated with the subject's TTCT scores before, during, and after a time period that will include creativity training.

By creating a psychometrically robust creativity assessment measure as a companion to the TTCT, we will be able to create an assessment tool that can map creativity training to development and real world practice to impact. These findings will help guide instructional content and training exercises for creativity training/ intervention in a large number of settings. Increasing creative capacity in individuals goes beyond classroom methodology to applied execution in real world scenarios for impactful change. The DTCT has the potential to become a new industry assessment norm for business professionals, educators and executives across all disciplines and industries as the creativity assessment that matches problem solving needs of the twenty-first century.

7.1 Wider DTCT Administration and Further Refinement

In order to reduce potential confounding covariates from the original prototype and increase statistical robustness, we will administer the DTCT to a larger sample of subjects for this study. We will correlate quality and quantity long-term data from the subjects that participated in the creativity capacity building project, namely information from both neuroimaging and creativity assessments will be analyzed. We will administer the DTCT to two-three additional graduate-level courses at Stanford's d.school to further refine the test language, to create a test administration manual, and to strengthen the establishment of a norm group for reliability and validity. The results from the wider administration of the DTCT will also help us determine the effectiveness of the case-driven hybrid format as real-world scenario.

References

Ball OE, Torrance EP (1984) Torrance tests of creative thinking streamlined scoring workbook, figural A. Scholastic Testing Service, Bensenville, IL

Barron F, Welsh GS (1952) Artistic perception as a possible factor in personality style: its measurement by a figure preference test. J Psychol 33(2):199–203. doi:10.1080/00223980.1952.9712830

Bott N, Quintin E-M, Saggar M, Kienitz E, Royalty A, Hong DW-C et al (2014) Creativity training enhances goal-directed attention and information processing. Think Skills Creat 13:120–128. doi:10.1016/j.tsc.2014.03.005

Carson SH, Peterson JB, Higgins DM (2005) Reliability, validity, and factor structure of the creative achievement questionnaire. Creat Res J 17(1):37–50. doi:10.1207/s15326934crj1701_4

Dietrich A, Kanso R (2010) A review of EEG, ERP, and neuroimaging studies of creativity and insight. Psychol Bull 136(5):822–848. doi:10.1037/a0019749

Dow GT, Mayer RE (2004) Teaching students to solve insight problems: evidence for domain specificity in creativity training. Creat Res J 16(4):389–398. doi:10.1080/10400410409534550

Goel V (2014) Creative brains: designing in the real world. Front Hum Neurosci 8:241. doi:10.3389/fnhum.2014.00241
Gough HG (1979) A creative personality scale for the adjective check list. J Pers Soc Psychol 37:1398–1405
Guilford JP (1950) Creativity. Am Psychol 5(9):444–454
Hawthorne G, Quintin E-M, Saggar M, Bott N, Keinitz E, Liu N (2013) Impact and sustainability of creative capacity building: the cognitive, behavioral, and neural correlates of increasing creative capacity. Springer, Cham, pp 65–77. doi:10.1007/978-3-319-01303-9_5
Katona G (1940) Organizing and memorizing: Studies in the psychology of learning and teaching. Morningside Heights, Columbia University Press, New York
Kienitz E, Quintin E-M, Saggar M, Bott NT, Royalty A, Hong DW-C et al (2014) Targeted intervention to increase creative capacity and performance: a randomized controlled pilot study. Think Skills Creat 13:57–66
Mednick S (1962) The associative basis of the creative process. Psychol Rev 69(3):220. doi:10.1037/h0048850
Mendelsohn GA (1976) Associative and attentional processes in creative performance1. J Pers 44 (2):341–369. doi:10.1111/j.1467-6494.1976.tb00127.x
Ridley DR (1979) Barron–Welsh scores and creativity: a second look. Percept Mot Skills 49(3):756–8
Royalty A, Lindsay NO, Bernard R (2014) Acting with creative confidence: developing a creative agency assessment tool. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer International Publishing, Heidelberg, pp 79–96
Runco MA (2004) Creativity. Annu Rev Psychol 55:657–687. doi:10.1146/annurev.psych.55.090902.141502
Runco MA, Plucker JA, Lim W (2001) Development and psychometric integrity of a measure of ideational behavior. Creat Res J 13(3–4):393–400. doi:10.1207/S15326934CRJ1334_16
Saggar M, Hawthorne G, Quintin E-M, Kienitz E, Bott NT, Hong D et al (2014) Developing novel methods to assess long-term sustainability of creative capacity building and applied creativity. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer International Publishing, Cham, pp 29–39, doi:10.1007/978-3-319-06823-7_3
Schaefer CE, Bridges CI (1970) Development of a creativity attitude survey for children. Percept Mot Skills 31(3):861–862. doi:10.2466/pms.1970.31.3.861
Torrance EP (1990) Torrance tests of creative thinking. Figural forms A and B. Scholastic Testing Service, Bensenville, IL
Wallach MA, Kogan N (1965) Modes of thinking in young children. Holt, Rinehart & Winston, New York
Zucker S (2003) Fundamentals of standardized testing. Pearson Inc, New York

Innovation in Creative Environments: Understanding and Measuring the Influence of Spatial Effects on Design Thinking-Teams

Claudia Nicolai, Marie Klooker, Dora Panayotova, Daniela Hüsam, and Ulrich Weinberg

Abstract Design thinking is a methodological approach addressing to solve wicked problems with multidisciplinary and cross-functional innovation teams. A flexible work environment enables innovation teams using this approach to ideate, create and design. This paper investigates the impact of creative environments on team wellbeing and performance as an outcome of perceptions, feelings and interactions of individual team members with their respective environment. The study introduces a new qualitative method, cultural probes, as an empirical instrument. The results of this study indicate that innovation teams need access to flexible spatial environments to fulfill their innovation process tasks (performance-oriented perspective), but also need different working zones to foster their team-wellbeing related activities (team-oriented perspective).

1 Creative Environments as Enabler of Innovation Processes

The ability to be innovative is one of the most important survival determinants of a company (cf. Zollenkop 2006, p. 10). Due to the increasingly dynamic global environment with its innovation pressure, companies are looking for new ways to promote creativity, inventions, ideas and new forms of innovation management within the company (cf. Herbig et al. 2008, p. 94). Previous research has focused on the process of innovation, its methods and tools (cf. Müller-Prothmann and Dörr 2009). In particular, design thinking is being promoted as a critical practice (Messner et al. 2008) in fostering innovation, particularly through collaborative processes of learning and knowledge creation (Martin 2009; Starkey and Tempest

C. Nicolai (✉) • M. Klooker • D. Panayotova • D. Hüsam • U. Weinberg
HPI School of Design Thinking, Hasso Plattner Institute for Software Systems Engineering, Campus Griebnitzsee, 14482 Potsdam, Germany
e-mail: claudia.nicolai@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_9

2009; Dunne and Martin 2006; Dym et al. 2006). Overall it represents a mind-set including a repertoire of tools, methods and techniques that can support a profound shift in organizational problem solving (Burry 2005; Dunne and Martin 2006). Organizations have started to adopt some of the specific problem solving practices, methods and tools, that have been practiced by designers in dealing with framing and re-framing complex design challenges.

Recent publications show the growing consideration of designing creative office environments (e.g. Groves and Knight 2010). The increasing occupation with spatial factors in the daily routine of individual office workers and employees is largely supported from two sides: The technical side enables increasing work digitalization and mobility. On the scientific side there are studies showing that productivity increases of up to 50 % can be achieved for office workers if spatial factors are considered (Walden 2011). Research results in neurology cast new light on the way we learn and work. Since industrialization, we have become used to environments for learning and working that do not suit the capacities we have developed over the course of evolution. John Medina goes so far as to say: "Anyone wanting to create a professional environment that diametrically opposes the special capacities of the brain will probably end up with a small office" (Medina 2014, p. XV).

However, there is a need to establish effective workspaces in order to enable appropriate spatial interactions as well as to create physical environments for diverse innovation activities (Moultrie et al. 2007, p. 53). Although there seems to be a consensus that the office environment has an effect on productivity of the individual employee, guidelines and methods for setting up new spaces, which suit in particular the specific needs of teamwork, its culture and objectives, are missing. According to Thoring and Müller (2011) a flexible working environment enables and supports the ability of teams to innovate, create and design. Peters (1992, p. 413) depicts space management as probably one of the most ignored and simultaneously the most powerful tool for the implementation of cultural change within organizations to foster innovation and learning within organizations. Work environments for creativity are often designed rather intuitively or based on often-cited examples of best practices (e.g. IDEO in the 2000s, Google nowadays). For example Magadley and Birdi (2009, p. 315) discuss the advantages of innovation labs as "dedicated physical environments or facilities with collaborative work spaces in which groups and teams of employees can engage with each other in order to explore and extend their thinking beyond and above normal boundaries". They suggest that spaces designed for fostering creativity have a different physical layout, a range of high- and low-tech supporting tools, and expert facilitators (see Moultrie et al. 2007) and show that such spaces have a positive impact on creativity and on participants' attitudes towards it. However, "the research studies on the spatial support of creative thinking with reference to the arrangement of a creativity-promoting work environment have not yet resulted in any generally valid, integrated complete solution" (cf. Rieck 2011, p. 43). Only insufficient and rather fragmented academic research has been conducted on the attributes of effective physical environments supporting creativity and innovation of teams.

Yet there is limited scientific proof of the benefits or the wider effects of the work environment on team performance and team wellbeing.

2 Understanding the Influence of Creative Environments: Cultural Probes as Research Method

In order to better understand the influence of spatial factors on innovation teams with regard to team performance and team wellbeing new methodological approaches are needed. The method of cultural probes was chosen as an appropriate mean for the elicitation of hidden dimensions empowering and preventing innovation teams to leverage their creative potential. The method of cultural probes was presented first by Gaver, Dunne, and Pacenti in 1999 and defined to be "a design-led approach to understanding users that stressed empathy and engagement" (Gaver et al. 2004, p. 53). Applying Cultural Probes generally aims at gaining insight into the daily lives and environments of the group under study. The probes are small packages that can include any sort of artifact (like a map, postcard, camera or diary) along with evocative tasks, which are given to participants to allow them to record specific events, feelings or interactions. The aim is to elicit inspirational responses from people over time, in order to better understand their culture, thoughts and values as well as their activities better (Gaver et al. 1999). Until now, they have become a common method in design research and are also frequently used in scientific research as a means for gaining qualitative data.

The use of cultural probes as a method is a mean to produce informational and comprehensible results, to identify concepts, constructs and hypotheses. Therefore it is often used in research settings that are aiming to generate conceptual frameworks. In comparison to common ethnographic research, the advantage of this method is that the researcher does not have to be present while data is collected. This enables not only the accumulation of a larger amount but also a bigger diversity of data over time. It also encourages the test person to share self-reports with the researcher and to provide more personal information than in presence of an interviewer or an observer. The cultural probes' tasks are carried out by participants in their own pace and away from the researcher's interference.

Nevertheless, the method of cultural probes has some shortcomings. Gaver et al. (2004, p. 54) himself calls cultural probes a "purposely uncontrolled and uncontrollable approach" and highlights the difficulty of measuring validity within the data collected. According to him they form "an approach that values uncertainty, play, exploration, and subjective interpretation as ways of dealing with those limits" (Gaver et al. 2004, p. 53). As a qualitative research instrument which includes different data collection tools, varying from inspirational material to open-ended tasks, only qualitative data analysis techniques (e.g. content analysis) can be performed.

3 Designing a Cultural Probes-Set

Bearing in mind the discussion about the use of cultural probes, we applied the method accordingly and designed the probes primarily to provoke inspirational responses from members of innovation teams concerning their perception of space. The cultural probes are used to identify beneficial and hindering spaces and spatial attributes within the context of the innovation space. We decided to additionally conduct qualitative interviews with the participants after the return of the cultural probes. This helped us to better understand their responses and offered an opportunity for further explanations. Through this we were able to conduct a qualitative content analysis of the collected data that led to an insightful outcome that will provide the basis for further experimental research to be conducted in the future.

The cultural probes tool-kit was designed to investigate the influences that the spatial environment might have on creativity in innovation processes and wellbeing in design thinking-teams. It contained three different tasks that helped us to understand the perception and the influence that spatial factors might have during a complete innovation project. Figure 1 shows the design of the cultural probes tool-kit highlighting its rather playful nature.

The cultural probes entailed three parts that were to be completed over a period of 3 weeks. To suggest an order of completion the tasks were put in three separate envelopes and numbered. Yet the participants were free to work on the tasks at any time during their innovation project. Each part was estimated to take at least 45 min. In addition to the data collected through this method, short qualitative interviews were conducted with the participants when the boxes were returned to the research term.

The tasks were designed to measure different aspects. Task 1 aimed at documenting the use of different spatial settings throughout the whole innovation

Fig. 1 Cultural probes tool-kit

process of the team. Additionally team members were asked to indicate their individual and the team's emotional state when working in different locations. Task 2 focused on the design of specific spaces and its impact on fostering specific working modes in terms of team wellbeing: teamwork (creation, ideation and prototyping), collaboration (knowledge, learning, and reflection), presentation (share and display project outcomes), as well as communication (socializing and relaxation). Task 3 was rather visionary and asked the participants to create their personal dream space.

To motivate the participants over the 3 weeks of time, incentives, such as a chocolate bar, were added. Additionally practical items (i.e. pens, glue, markers, scissors etc.) were added. All these tools were of high quality. As further incentive for participating the students were free to keep all material that was not used to complete the tasks. The cultural probes were wrapped in a cardboard box with the size of a DIN A 4 folder, to be carried home and brought back easily.

We selected different student teams of HPI School of Design Thinking at Potsdam as the study group. The sample size consisted of three different innovation teams working on a design project over a period of 8 weeks. The sample size consisted of 12 participants in total. This group was chosen due to its diverse nature, which results from the multidisciplinary background of design thinking teams. Due to the fact that the students together formed three teams, it was possible to compare their individual perception with the perception of other team members as well. All participants were moderately experienced in design thinking and went through a 6 weeks-training course before starting their innovation project.

4 Measuring the Perception of Spatial Effects

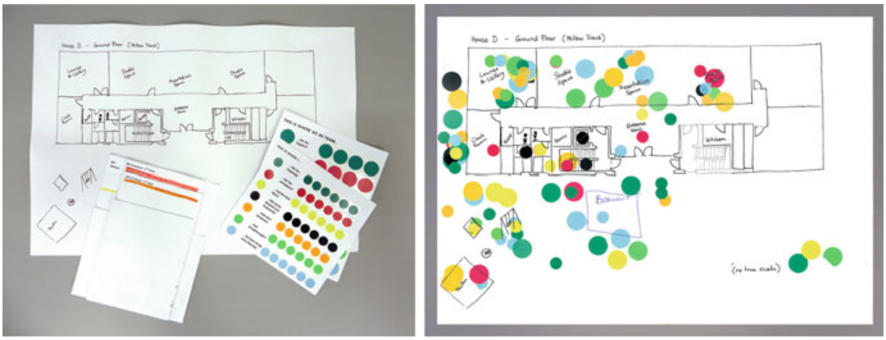
4.1 Space as a Medium and Mediator of Innovation Team Performance and Team Wellbeing

Previous work by Gensler (2008, 2013), Groves and Knight (2010), and Doorley and Witthoft (2012) suggests that performance and wellbeing of innovation teams could be improved when different spatial settings are used for distinctive tasks.

The HPI D-School offers a variety of spaces that either promotes teamwork, meeting spaces, lounge spaces, workshop spaces, reading and silent spaces as well as spaces for individual work. In this research setting each team was assigned to their own team space in an open studio setting shared with eight teams in total. Each team space is typically equipped with moveable whiteboards, one flexible table, high chairs and material for teamwork. Spaces commonly used by all teams are the presentation space equipped with various projectors and flexible seating (including sitting cubes, chairs and sofas). The lounge as well as the kitchen is the assigned communal space (including various sofas, bean bags, and sitting cubes) whereas the library is used as a reading room and a room for silent work (for individuals). The workshop spaces (one for digital prototyping and one for handicraft work) can be found on different floors and offer facilities for teams as well. If weather conditions are permitting it, teams can also use the patio in front of the building and a teamwork pavilion (designed to host a maximum of two teams). Because different spaces can foster different work modes the teams have been encouraged to change their workspace according to their innovation project and team wellbeing needs.

The analysis of the usage patterns of the three teams under scrutiny revealed that the design thinking-teams have used all different spaces during their project. Figure 2 shows an example of a team with dots indicating the usage and emotional states associated with different spatial settings during the whole innovation process. All three teams displayed the same usage patterns in terms of time spent at the different locations. It is interesting to notice that the teams quite often moved to a different space when they felt stuck in the innovation process as one interviewee underlined: “We definitely had great ideas here [in the team space] but most of the breakthrough moments happened when we changed to another space.” And another person said: “What I did with my team was to push them out of the space where we were currently working when we got stuck. .... When we entered a new space, outside or inside, we felt unblocked by the surrounding itself.”

These findings indicate that team performance can be fostered by the change of the team’s workspace location. Overall, we discovered that the teams did not only change the location of their teamwork based on the stage of the innovation process stage they were currently working on; e.g. moving to the workshop space to build prototypes, moving to the communal space to relax, or finding a quiet space in the library to reflect—individually—on their team performance. They quite often changed their spatial setting as a mean to ignite their creative potential and therefore team performance. Quite often the teams went to seek a space that seemingly offered them the opposite wellbeing situation to the current state.



Floor plan of innovation studio and instruction to mark the different working zones and their impact on team performance and team wellbeing

Example of the perception of the different working zones indicated with stickers of one test team

Fig. 2 Example of the cultural probes task for team performance and team wellbeing

When the assigned team space was considered as too noisy and too distracting they moved outdoors or to the library. Thus, a lot of "misuses" of additional spaces took place over the time of their innovation project, e.g. using the silent space of the library as the place to perform an intense—and loud—round of ideation, working in the workshop space (in the basement) to find a—silent—space without any noise and other teams around them for individual "thought-work" without touching any of the tools.

In addition to the use of the various spaces for the distinct tasks of an innovation process we also asked all team members to share with us their assessment of their individual wellbeing and their perceived wellbeing of the whole team in different locations. Our interviews after the completion of the cultural probes revealed that the team members were able to indicate very accurately their own emotions about how they felt individually about working with their team in the different locations but were not able to express their team's overall feelings with the same accuracy. It is also worth noticing that we have got completely different answers from the members of the same team about their individual feelings about working in a specific space, in particular for their assigned team space and the outdoor spaces (either on the patio or the lawn in front of the building). For example, one team member stated that he felt energized "... when we went outside and walked around barefoot on the ground" whereas another member of the same team stated that he did not think it helped the team's progress because "... we lacked the tools to take all relevant information with us to our temporary workspace outside." Therefore it is no wonder that the assigned team space—inside—and the most often chosen change to a temporarily team space—outside—were the two most often marked spaces in our cultural probes set that received positive as well as negative connotations as working spaces. To quote two other interviewees: "In our assigned team space we felt imprisoned. ... Trying to create ideas was not much fun here but everywhere else." Another person stated: "This was OUR team space with a lot of emotions going on ... You should really explore what YOUR TEAM SPACE can do for you as a team."

When analyzing all individual emotional states associated with the use of specific spaces we found an interesting relationship with the frequency of usage. The spaces that were used most often are in the order of frequency: the assigned team space, outside spaces, presentation space, workshop spaces, lounge, and library. In particular the first two ones received positive as well as negative emotions and feelings: "What I personally found interesting is that our own team space it the place where we felt most confused." And another test person stated: "In the outside pavilion we had great moments but it also sometimes did not work out at all." The three latter ones received positive emotions as well as where spaces individual team members were indicating that the teamwork had breakthrough moments with regard to the team's performance.

Overall, positive emotions are most often experienced at places to socialize or to recreate (e.g. lounge and kitchen). Spaces to create and make (e.g. workshop spaces) as well as sharing spaces (e.g. presentation spaces) are also perceived as uplifting. Whereas at those spaces where most of the intense teamwork (concentration and convergent thinking) took place (e.g. team spaces and outdoor spaces) are often related with mixed emotions.

4.2 Different Working Modes of Innovation Teams Transported Through Spatial Environments

Design thinking has been promoted as a problem-solving activity concentrating on wicked or ill-defined problems (Buchanan 1992). As a problem-solving approach it represents a collaborative process of knowledge creation and iterative learning towards innovation. The different stages of this teamwork approach are described as a system of activities (Brown 2008). Definitions of these activities vary from three up to six different process stages.

Drawing from the experience of several innovation spaces as well as research on the proximity of team members and group effectiveness, it is evident that stimulating environments can foster innovation team performance (Beckman and Barry 2007; Doorley and Witthoft 2012; Moultrie et al. 2007). While existing research sees creative problem solving as a dynamic, complex and subjective process that takes place in a stimulating environment, we do not know much about how we create the conditions to flourish this process beyond fostering teamwork spaces. Besides the physical set-up of team workspaces, the atmosphere around and the adjacent spaces are relevant for team performance and wellbeing. Groves and Knight (2010) have assessed different creative environments and have identified three different working modes besides teamwork (collaboration) that enable knowledge creation as well: stimulation, play, and reflection. Gensler (2008, 2013) underlines the need for additional facilities enabling: socializing, sharing, connecting, and learning. Based on this previous research and experience at the HPI D-School, we have defined four spatial categories as being relevant besides collaborative spaces for teamwork throughout the design thinking process:

- Build and make
- Present and show
- Communicate and reflect
- Socialize and relax.

Acknowledging the subjectivity of preferences concerning workspaces (for individuals, for teams, and for larger groups) we were interested in the individual perception of the mentioned types of spaces. We asked the participants of our study to share with us what kind of atmosphere they prefer in these spaces. The cultural probes task included a written description of the space and its assigned activities and additional questions about the look, sound, smell of the place. Aiming at an inspirational format, we decided to pose the open-ended questions in the format of a postcard without suggesting a specific order for the answers (see Fig. 3). The participants were asked to create four different postcards; one for each spatial

Cultural probes task kit to visualize the individually preferred spatial environment for different working modes

Examples of illustrated postcards done by the participants of the study

Fig. 3 Example of the cultural probes task on spatial settings for different working modes

category. Furthermore, the participants were asked to record the surroundings and to take pictures of analogous places they experienced in real life and which they see as objects of reference for the specific working modes.

It is interesting to notice that in particular for the "*build and make*"-space the range of individual preferences is very broad. The individually mentioned preference for the surrounding sound can exemplify this: from silence to soft music without lyrics to loud electronic beats nearly everything was mentioned. This already indicates one of our key findings: individual preferences are not always compatible with each other. Nonetheless the content analysis led us identify two different patterns: The participants were distinguishing '*spaces for inspiration*' from '*space for doing*'. Inspirational spaces offer a lot of visual clues, different kinds of material at display, and different seating and working positions/working stations. This describes an informal atmosphere with enough space to walk around and to change positions. Some of the participants even referred to being in nature as their ideal inspirational space. Spaces for 'doing' are associated with a clean and tidy workshop space with access to all tools needed (for building as well as digital prototyping).

The associations and feelings about the "*present and show*"-space showed some differences as well. Overall the participants clearly distinguished between spaces and situations in which merely presenting—and not sharing—takes place (an atrium, a stage) from situation where both take place. For the first one, different participants distinguished between real and virtual presentations. For the latter they wish for a comfortable and relaxing situation for the audience and an opportunity to interact with them. Thus, rooms should not be distracting too much.

The space to "*communicate and reflect*" is obviously a space for the whole team and/or for the individual alone. The participants depicted different situations in which they were alone, in smaller groups of two to three people or in their whole team. It is interesting to notice that preferred spatial settings never included a larger group or audience. The ideal locations to practice this working mode range from the assigned teamwork space to a walk in the park (alone or with another team member). The participants mentioned that they need tools for this, like pens, post-it's or a notebook, screens for visualization, surfaces to write on, etc.

Spaces to "*relax and socialize*" were also associated with a broad variety of settings, activities and people. Besides the fact that all participants mentioned practicing sports, being with friends, eating and drinking, and spending time in nature the prominent overarching topic has been a surrounding trustful atmosphere. Thus, either the space itself or the time dedicated to it should provide the participants with the feeling that they are—more than—allowed, invited, to do so: relax and socialize.

Both activities, designing the postcards and doing the visual and audio recording, together composed a colorful picture that provided valuable insights about the individual preferences for the different additional spatial categories. In the postcards activity, the students were free to merge experiences of multiple spaces to create a catalogue of characteristics for each spatial category. However, this included a further layer of interpretation because individual preferences displayed a large variety of possible spatial solutions.

4.3 Designing Ideal Creative Environments for Individuals and Teams

The third part of the cultural probes set was merely visionary. Assuming that the tasks were completed in the assigned order we expected the participants to have gained awareness about their spatial surroundings and the topic of spaces for creativity in general. This created the mindset to design the creative space of their dreams. To fulfill this task, the participants were provided with material and tools. We decided to provide 36 samples of material for interior design including printings of surrounding elements such as clouds or water. The haptic experience of different materials supposedly helped the students to better imagine the use. Additionally the students were asked to add people-posture-cards that we provided. These additional 36 cards showed different sizes of groups and singular people in different working positions (e.g. standing, sitting, etc.). This gave us a hint of the activities assigned to the space and encouraged the students to keep these in mind when designing the space. Purposely no floor plans or pre-designed forms were provided. The spaces were to be created on a blank paper. On return of the boxes, we realized that a few even used the wrapping, hence the cardboard box and created a 3-D model of the space (see Fig. 4).

On average the participants have used 11 of the 36 provided material samples and 8 of the 36 different working positions. Figure 5 shows the most often used material samples.

Cultural probes task kit with material samples

Example of a creative space for innovation designed by one participant

Fig. 4 Example of the cultural probes task on designing creative spaces for individuals and teams

Numbers indicate the frequency of usage in our sample of n = 12 participants

Fig. 5 Example of the cultural probes task on designing creative spaces for individuals and teams. *Numbers* indicate the frequency of usage in our sample of n = 12 participants

When analyzing the aesthetics of the used materials we were able to identify two different material combinations: On the one hand we found the combination of rather more natural materials (plank floor, cork, grass carpet) for floors with a translucent atmosphere (frosted glass, wallpaper with the depiction of skies and water). On the other hand the combination of natural materials (plank and concrete floor) with a more industrialized, but bright atmosphere (different textured materials in white, grey and red, brick and white wall).

The individual designs of ideal creative spaces consisted typically of three different working zones. All of the 12 designs included a designated zone for team spaces. The two other zones are a recreation space (designed 9 times), presentation space (6 times), kitchen as lounge (5 times), silent space (5 times), make space (4 times), ideation space (3 times), or a private individual space (2 times).

The designs were quite often very minimal offering enough space for "... your body; you are in a movement while your mind is working". Some of the participants even mentioned the possibility of opening and maximizing or closing and minimizing the space according to their (team) needs at a certain time. This has been also one of the most recurring themes in the interviews about this cultural probes task. The freedom to be able to choose where to work, with whom to sit and talk seems to be a dominant design principle as one interviewee said: "You need to allow people to change the space. ... People need to express personality as part of the company. ... Encouraging change in your innovation space is the first step to change."

5 Discussion of Findings and Further Research Questions

Previous research in the field of creative environments has been focusing on the description of different physical characteristics of alternative environments for innovation teams. However, if doing so without considering the usage of the space by innovation teams and its impact on team wellbeing and performance will provide little insights into the role and effectiveness of place and place-making (Moultrie et al. 2007). The authors have also experienced many different examples of innovation spaces during their analysis of best practices, e.g. co-working space, innovation greenhouses and idea saunas.

Using the method of cultural probes we were able to define a research method that helps in particular to reveal the impact of creative environments on team and individual wellbeing over time. Overall, the method of cultural probes has been used to identify beneficial and hindering spaces and spatial attributes within the context of the creative space during an innovation project of teams, to discover different spatial requirements for diverse working modes outside the individual teamwork environment as well as to collect the designs of the ideal, imagined innovation spaces of the participants. It needs to be noticed that this has been a rather descriptive qualitative study and that the impact of creative environments on team performance is still a research question under scrutiny.

However, we were able to identify some key findings: Overall, different patterns could be detected concerning the emotions of innovation team members in different team working spaces—at different times. As the participating innovation teams were free to move around the HPI D-School building and use the spaces for different purposes besides only their assigned team-space, it was difficult to clearly define the impact of a space. For example the library was used for an ideation session, hence a creative act, while others used it to concentrate and reflect. Yet, it can be derived that switching spaces might trigger the switch from being a "stuck team" to a team seeking a new start for their flow of creativity with positive emotions and through this promotes creativity. The collected data from the second task of the cultural probes set proved that preferences concerning the set-ups for different working modes in the design thinking process are very subjective as they differed greatly. For some, listening to loud and fast music seems to be a means to recreate, while others prefer a silent surrounding. For some participants open and crowded spaces are fostering creativity while others like clean and separated spaces. The third task confirmed the aspect of subjective preferences. Innovation teams need different working zones; not only one assigned team space. All participants completed the task to imagine the creative space of their dream by designing a compilation of several spaces that were assigned to different tasks or functions. In particular zones for sharing, connecting and learning seem to be important for team wellbeing. A reduced design combining rather natural materials with a bright atmosphere and a touch of industrial aesthetics with steel, bricks and a lot of white in general seem to be a common wish for creative spaces.

It would be interesting to further investigate if the above-mentioned preferences for outdoor (innovation and recreation) activities as breakout are a phenomenon due to the fact that the study has been conducted over the summer in Germany. We suppose that this depicts another important aspect to consider in the configuration of innovation spaces: the design and usage of team workspaces and individual workplaces as well as the ability to change between different, physically active, loud, and visual vibrating team-oriented working phases, and quiet, thoughtful and reflective individual-oriented working phases. The expertise of one author being involved in the design and setup of a D-School in Malaysia underlines this conclusion. Local weather conditions (high temperatures, humidity, and rain) prevent working outside and breakout-spaces for individuals had to be—and could be—created indoor as well.

Bearing in mind all the above, it can be concluded, that subjectivity in aesthetics plays an important role once it comes to the design of innovation spaces. Nonetheless, what all participants have in common is the need for different creative spaces and working zones fitting the different tasks and activities throughout an innovative team process.

(Mental) freedom is needed to adjust the workspace according to, on the one hand, the tasks to be done at a certain point in the process, on the other hand, to set up the space to meet individual (team's) preferences. This accounts to the model of Freiraum, that Auernhammer and Hall (2013) defined to be promoting creativity. Furthermore, it seems that creativity was fostered when teams were encouraged to select a specific space to work at and then transformed it to make it their own. The flow of emotions went up when teams moved to another work zone and had to transform it—temporarily—to make it their own.

Overall the empirical study, as described above, showed that having the freedom to choose is a promoting factor for creativity. Yet, creating a space that is a blank page, does not seem to help. A space that fosters creativity rather suggests a direction of usage but also leaves room to change it.

This conclusion opens up two fields for future research: The first topic to be addressed is the academic rigor given to the relationship between creative spatial settings and the perception and evaluation of team wellbeing and performance. In a follow-up study we would like to analyze team processes in-situ to add insights to our retrograde approach with cultural probes. The second topic encompasses the innovation team process itself. We would like to better understand how to enable innovation teams to make the choices that best possibly promote their creativity. This action-based research approach will shed further light on creating and using the innovation space due to the team's changing tasks and activities throughout the process.

References

Auernhammer J, Hall H (2013) Organisational culture in knowledge creation, creativity and innovation: towards the Freiraum model. J Inf Sci 39(1):1–14

Beckman SL, Barry M (2007) Innovation as a learning process: embedding design thinking. Calif Manage Rev 50(1):25–56

Buchanan R (1992) Wicked problems in design thinking. Des Issues 8(2):5–21

Brown T (2008) Design thinking. Harv Bus Rev 86(June):84–92

Burry MC (2005) Homo faber. Archit Des 75(4):30–37

Doorley S, Witthoft S (2012) Make space: how to set the stage for creative collaboration. Wiley, Hoboken, NJ

Dunne D, Martin RL (2006) Design thinking and how it will change management education: an interview and discussion. Acad Manage Learn Educ 5:512–23

Dym CL, Agogino AM, Eris O, Frey DD, Leifer LJ (2006) Engineering design thinking, teaching, and learning. IEEE Eng Manage Rev 34(1):65–92

Gaver WW, Dunne A, Pacenti E (1999) Cultural probes. Interactions 6(1):21–29

Gaver WW, Boucher A, Pennington S, Walker B (2004) Cultural probes and the value of uncertainty. Interactions 11(5):53–56

Gensler (2008) 2008 Workplace survey—United States. Washington, DC

Gensler (2013) Design matters! 2013 U.S. workplace survey. Washington, DC

Groves K, Knight W (2010) I wished I worked there! A look inside the most creative spaces in business. Wiley, Hoboken, NJ

Herbig B, Glaser J, Gunkel J (2008) Kreativität und Gesundheit im Arbeitsprozess: Bedingungen für eine kreativitätsförderliche Arbeitsgestaltung im Wirtschaftsleben. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund

Magadley W, Birdi K (2009) Innovation labs: an examination into the use of physical spaces to enhance organizational creativity. Creat Innov Manage 18(4):315–325

Martin RL (2009) The design of business: why design thinking is the next competitive advantage. Harvard Business School Press, Boston, MA

Medina JJ (2014) Brain rules, 2nd edn. Pear Press, Seattle, WA

Messner M, Kornberger M, Clegg S (2008) Critical practices in organizations. J Manage Inq 17(2):68–82

Moultrie J, Nilsson M, Dissel M, Haner U, Janssen S, Van der Lugt R (2007) Innovation spaces: towards a framework for understanding the role of the physical environment in innovation. Creat Innov Manage 16(1):53–65

Müller-Prothmann T, Dörr N (2009) Innovationsmanagement: Strategien, Methoden und Werkzeuge für systemische Innovationsprozesse. Hanser, München

Peters T (1992) Liberation management: necessary disorganization for the nanosecond nineties. Knopf, London

Rieck A (2011) Beitrag zur Gestaltung von Arbeitsumgebungen für die Wissensarbeit. Jost Jetter, Heimsheim

Starkey K, Tempest S (2009) The winter of our discontent: the design challenge for business schools. Acad Manage Learn Educ 8(4):576–586

Thoring K, Müller R (2011) Understanding the creative mechanisms of design thinking: an evolutionary approach. In: Proceedings of the second conference on creativity and innovation in design, pp 137–147

Walden R (2011) Auch mal im Garten arbeiten. In: Deutsches Architektenblatt (08/2011), pp 14–17

Zollenkop M (2006) Geschäftsmodellinnovationen. Deutscher Universitäts, Wiesbaden

Building Blocks of the Maker Movement: Modularity Enhances Creative Confidence During Prototyping

Joel Sadler, Lauren Shluzas, Paulo Blikstein, and Riitta Katila

Abstract Can we enable anyone to create anything? The prototyping tools of a rising Maker Movement are enabling the next generation of artists, designers, educators, and engineers to bootstrap from napkin sketch to functional prototype. However for technical novices, the process of including electronic components in prototypes can hamper the creative process with technical details. Software and electronic modules can reduce the amount of work a designer must perform in order to express an idea, by condensing the number of choices into a physical and cognitive "chunk." What are the core building blocks that might make up electronics toolkits of the future, and what are the key affordances? We present the idea that modularity, the ability to freely recombine elements, is a key affordance for novice prototyping with electronics. We present the results of a creative prototyping experiment ($N = 86$) that explores how tool modularity influences the creative design process. Using a browser-based crowd platform (Amazon's Mechanical Turk), participants created electric "creature circuits" with LEDs in a virtual prototyping environment. We found that increasing the modularity of LED components (i) increased the quantity of prototypes created by study participants; and (ii) increased participants' degree of perceived self-efficacy, self-reported creative feeling, and cognitive flow. The results highlight the importance of tool modularity in creative prototyping.

J. Sadler (✉) • L. Shluzas
Mechanical Engineering Department, Stanford University, Stanford, USA
e-mail: jsadler@stanford.edu

P. Blikstein
School of Education and Computer Science Department, Stanford University, Stanford, USA

R. Katila
Managment Science and Engineering Department, Stanford University, Stanford, USA

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_10

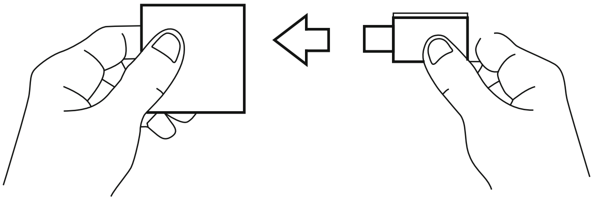


Fig. 1 Plug-and-play interaction with electronics modules

1 Introduction: Modules for Makers

Can we enable anyone to create anything? Maker culture encourages those with ideas to empower themselves with the *creative tools* to make their ideas tangible. In the world of electronic devices, the rising availability of novice microcontroller platforms, desktop physical fabrication tools, affordable electronic components, and a global online community, has fuelled more designers to *hack, mash* and *glue* their way to working electronic prototypes (Hartmann et al. 2008). Hundreds of DIY fabrication facilities, *Fablabs*, have emerged over the years to provide public spaces for digital-physical prototyping, ranging from laser cutters, and computer-controlled milling machines, and 3D printers (Anderson 2014). Tools that were once only accessible to large companies, or required professional knowledge, are now within reach of ordinary citizens. In the electrical domain, novices can now chose from a growing array of premade electronics modules, such as sensors and actuators (Fig. 1), that hide much of the technical detail into plug-and-play LEGO-like building blocks (Sadler et al. 2015a, b). These emerging tools have strong implications for Design Thinkers who wish to transform low-resolution prototypes from paper and duct-tape, into functional interactive devices. But what are the core building blocks that might make up electronics toolkits of the future, and what are their key affordances? Here we propose the idea that *modularity*, the ability to freely recombine elements, is a critical affordance for novice prototyping with electronics. What are the creative trade-offs with more modular electronics blocks? Do these blocks help or hinder the creative design process?

1.1 Modularity and the Black Box

When faced with a creative challenge, designers are encouraged to create many low-resolution solutions in a limited time period. These prototypes tend to highlight the essence of an idea, rather than the technical details of implementation. However prototyping with software and electronics can slow down the creative process by (i) requiring prior technical knowledge and (ii) introducing potential for errors in implementation. By using pre-defined modules, a designer may add blocks that encapsulate a desired functionality, while hiding the technical details. Here we define modules as encapsulated blocks of functionality that can be added to or removed from a system independently (Baldwin and Clark 2000).

A component with a high degree of modularity has fewer dependencies on outside variables. In prototyping, this implies that modules enable designers to freely try combinations of parts, much like adding bricks in a toy construction kit. For example if the designer decides to add a green LED light to a prototype, a plug-and-play interaction is a more modular interaction than building a light circuit from basic components. We can think of the light module as a "black box" that does not reveal its inner details, but rather emphasizes the core functionality. For designers who are trying to create novel solutions, the idea of "black box" that cannot be modified on the inside seems counter to the idea of creative freedom. On the other hand, the "black box" approach allows the designer to focus on the high-level design decisions such as, "Should we include a light?" before diving deeper into technical requirements such as, "What value of resistance should we include to limit the current?" Prototyping with black-box modules therefore has tradeoffs between flexibility and ease of use. The goal of this research is to better understand the effects of these modular tradeoffs on the designer's creative ability.

2 Background

2.1 The Modularity Tradeoff: An LED Example

What does high or low modularity look like with electronic components? Consider the following example of two multi-color light emitting diode (LED) components shown in Fig. 2. In the example of "low modularity," a low-cost tri-color LED can be added to a programmable microcontroller by connecting four individual pins, and choosing 3 appropriately sized resistors. In this case, the LED does not contain all of the components required to function reliably, and a designer must choose the correct resistors from thousands of possible resistance values.

Making an incorrect connection, or choosing a resistor that is too small or too large, results in a non-functional light. In the "high modularity" example, a self-contained

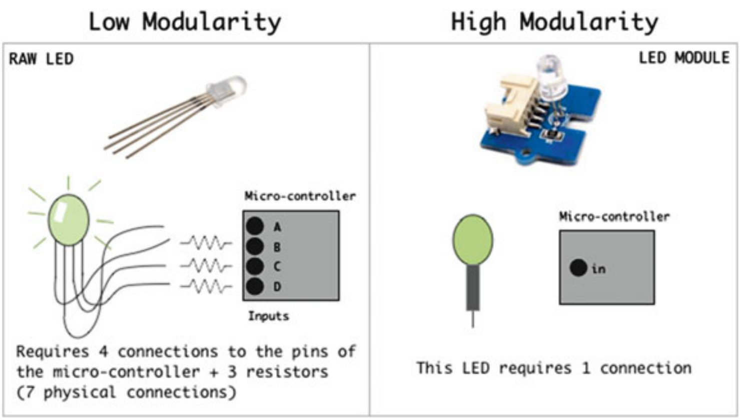


Fig. 2 Comparing the modularity of two LED components

Table 1 Comparing modularity trade-offs (+++ is best)

Metric	Less modular	More modular
Number of steps (to combine)	*More steps required*	+++
Difficulty threshold	*Increased initial difficulty*	+++
Error probability (success variability)	*Increased chance of error*	+++
Component cost	+++	*More expensive*
Technical learning	+++	*Reduced technical learning*
Flexible functionality	+++	*Harder to modify*

LED module contains the necessary components to function properly and a single connection to the outside world. In this case, an expert has made the resistor choices in advance, and the module has a simplified single connection interface. This module reduces the outside variables that may lead to errors, such as choosing bad resistors or plugging in the component in an opposite polarity. These contrasting examples illustrate that a key difference in prototyping with modules is that they tend to reduce the variability, difficulty, and number of steps required to combine components, in exchange for reduced flexibility and increased cost. These trade-offs are summarized in Table 1. Modules may be a poor choice if one's goal is to educate a designer on the technical aspects of a system, since the details are hidden. However, if the primary goal is to enable functional prototyping of a creative idea as quickly as possible, then modules are effective candidates. Modules are therefore the "chunks" that enable a designer to incorporate a whole unit of functionality; in much the same way that cognitive chunking (Miller 1956) enables efficient clustering of complex information. Here the authors use the concept of a "chunk" as equivalent to a module, as proposed by the design decomposition work of Pimmler and Eppinger (1994).

2.2 Prototyping Metrics

During prototyping we assume that an ideal creative toolkit will enable a designer to confidently create many prototypes in a given time frame, allow diverse ideas to be expressed, and place a designer in "the zone" of continuous cognitive flow. Since modularity tends to reduce the difficultly and number of steps required to create a prototype, we hypothesize that increasing the modularity of prototyping tools will have positive effects on:

1. **Prototype Quantity**: The number of distinct prototypes created in a given time period.
2. **Designer's Creative Feeling**: The degree to which designers feel they are in a creative state.
3. **Self-Efficacy**: The confidence that a designer has in his or her ability to create prototypes. (Csikszentmihalyi 1992).
4. **Cognitive Flow**: The degree to which a designer feels that a task's difficulty level matches one's perceived abilities (Csikszentmihalyi 1992).

The above metrics have interconnected effects on one another. For instance, prior research has shown that designing several prototypes in parallel can contribute to an increase in both the quantity and creativity of prototypes produced (Dow et al. 2010). This research focuses on the above metrics as general indicators of a desirable prototyping experience.

2.2.1 Bandura's Self-efficacy and the Confidence to Create

Creating prototypes with technology can be difficult for a technical novice. A designer who lacks a technical background in electronics and sensors, for example, may be less confident in his or her ability to make working prototypes that include electronic components. Bandura (1977) refers to the measure of a person's confidence in his or her ability to achieve a task as *self-efficacy*. In this research we consider a specific aspect of self-efficacy, referred to as *creative self-efficacy* (Tierney and Farmer 2002). For a prototyping task, we aim to understand how the qualities of a tool will encourage designers to feel confident in their ability to create. Prior work on the use of novice electronic prototyping toolkits found that using modular building blocks can be an effective way to encourage more fluid prototyping (Hartmann et al. 2005). The core premise of Hartmann's work is that modularity hides some of the technical details and reduces the perceived and actual difficulty to play with an unknown component. However, to the authors' knowledge there is limited prior data on the cognitive effects of modularity during creative prototyping.

2.2.2 Cognitive Flow and Modularity

Designers often describe a creative episode as a sustained burst of focused energy on the task at hand, where time fades into the background and one gets into *the zone*. The cognitive psychologist, Mihaly Csikszentmihalyi (1992), describes this state as a *flow state,* and his work has shown that being in a state of flow is positively correlated with creative performance. Csikszentmihalyi's work describes the typical conditions required to trigger a flow state, and found that flow is modulated by a balance between one's perceived skill and the perceived difficultly of a task. If a task is too challenging for a person's current skill level, Csikszentmihalyi's model predicts a state of anxiety. Similarly, if the challenge of a task is low and the perceived skill is high, the person may be in a state of low arousal or boredom. The flow state is characterized by a balance of challenge and skill, in a *channel of flow,* which is linked with higher task performance and creativity (per Fig. 3). This model has been validated by many researchers whose work illustrates the positive relationship between flow and task performance (Engeser and Rheinberg 2008). From Hartmann et al.'s work we see that modularity changes the prototyping experience by promoting components as "chunks" rather than as individual technical details. However, current research has yet to measure modularity's role in inducing a flow state during a design activity.

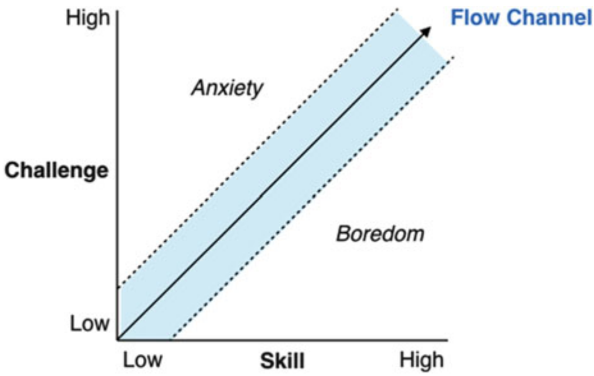


Fig. 3 Flow as a balance between skills and challenge (Csikszentmihalyi 1992)

3 Related Work

In order to examine the effects of tool modularity on the creative design process, we need a repeatable and modifiable task to serve as a basis for measuring prototyping performance. The methods we chose were modeled on two prior studies, one examining conformity in creative generation tasks (Marsh et al. 1996), and another by Kulkarni et al. (2014) that uses a crowd-sourced method to explore a creative prototyping task.

3.1 Conformity in Creative Generation Design Tasks

Marsh et al. (1996) illustrated that presenting design examples to participants prior to a generative design task influences one's creative output. Participants were shown pre-made examples of alien creatures and then asked to draw as many unique creatures as possible in a given period of time. The study found that showing design examples in advance increased the degree of conformance among design ideas. We consider this study as a strong model of a creative prototyping task, with a controlled manipulation. Specifically, we see a parallel between electronic modules and the pre-made design examples in the study by Marsh et al., since modules can be considered as pre-made units of function that a designer is provided in advance and may choose to include in a prototype.

3.2 Timing Effects on Creative Output: A Crowd-Sourced Design Task

In building on the prior study, Kulkarni et al. (2014) examined how creative output is affected by the timing when design examples are shown to study participants. Their research found that early and repeated exposure to examples increased creative design performance, as measured by the number of uncommon and novel features created by study participants. There was a correlation between exposure to examples and conformance, but this did not reduce the number of unique features incorporated into ideas. Most notably, this study showed the feasibility of using a crowd sourced web-based platform to recruit participants (N = 81) and collect a large number of prototype alien ideas using Amazon's Mechanical Turk. The use of a software environment to test creative prototyping is advantageous since it allows controlled modification of the environment. Our work builds on these two studies, in terms of "imagining alien creatures" as a creative prototyping task, but adapts the task to the creation of "electric alien circuit boards" with colored LED modules. To the authors' knowledge, no prior work has specifically examined the effect of modularity on creative output, self-efficacy, and cognitive flow.

4 Methods

Our research aims to understand how modularity affects creative prototyping from both quantitative and qualitative perspectives. For a controlled design task, we hypothesize that:

H1 Increasing the modularity of a prototyping tool enables designers to generate a higher quantity of prototypes in a given period of time.

H2 Increasing the modularity of a prototyping tool increases a designer's creative feeling, perceived self-efficacy, and cognitive flow. [Conversely, increasing a designer's exposure to technical details while prototyping (e.g., through the number of wiring options) decreases these cognitive measures.]

To test these hypotheses, we provided a generative prototyping challenge to a broad group of participants, and varied the degree of tool modularity for each participant. Based on a modified version of the prototyping task by Kulkarni et al., we asked participants to create as many unique and novel electric alien creatures as possible in a 10-min period (Fig. 4). We provided the following prompt, adapted from Marsh et al.:

> Imagine a planet like earth existing somewhere else in another universe. It is currently uninhabited. Your task is to design new creatures to inhabit the planet. The creatures in this world are very special, since they all are made up of tiny electric blocks. With 10 minutes allotted, draw and describe as many new and different creatures of your own creative design as you are able. Duplications of creatures now extinct or living on the planet Earth are not permitted.

We developed an HTML/Javascript based circuit creation tool that allowed participants to drag and drop colored LED's or draw grey paint on a 16×16 circuit board (Fig. 5). Users could submit as many prototypes as they wished in the time allotted, as well as provide a verbal description of each prototype. Participants were

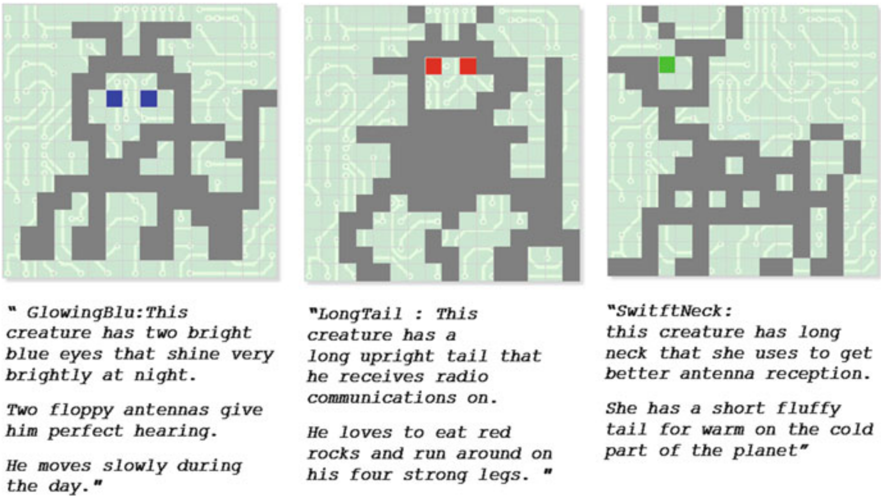


Fig. 4 Example electric creatures shown to the participants

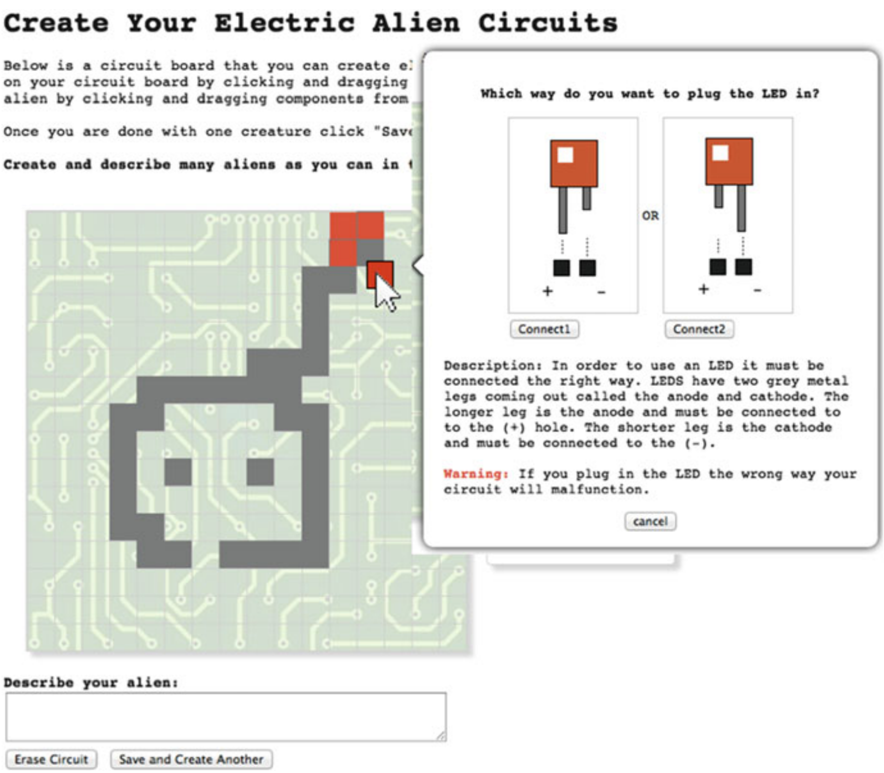


Fig. 5 The alien circuit interface where participants could drag LEDs onto a 16 × 16 grid

asked to physically arrange and describe each circuit board design, where each circuit would be a visual representation of a novel creature. The LED components were presented as a possible design option to add "virtual" colored lights. Using this software representation of a circuit allowed us to manipulate features of the interface for three user groups.

4.1 Participants and Groups

We recruited (N = 86) participants on Amazon's Mechanical Turk with a compensation of $1.00 USD per participant (48 male, 38 female, average age 36 years). Users were filtered to include only fluent English speakers located in the United States. Participants were randomly assigned to one of three groups. The control group was able to freely place LEDs on the circuit board as encapsulated modules, without interruption. When participants from Groups A or B attempted to place an LED on their circuit, they were interrupted by a technical description of LED polarity, and asked to plug in wire leads in the correct orientation before they would appear on the circuit (Fig. 6). Groups A and B have less modular interactions, as described in Sect. 2.1, since there is an additional choice that must be made before successfully adding each LED. Group B differed from Group A, in that they were presented with a randomized LED orientation and randomized plug orientation.

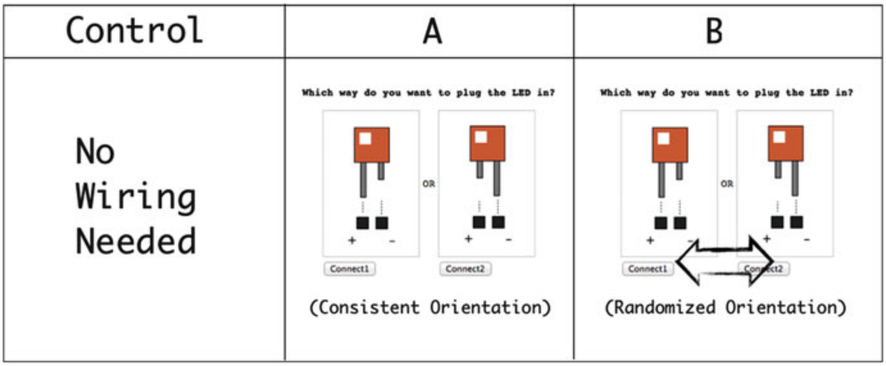


Fig. 6 Three groups. Groups A and B were given a technical description of LEDs. Participants had to select the correct wire orientation for the placement of each LED

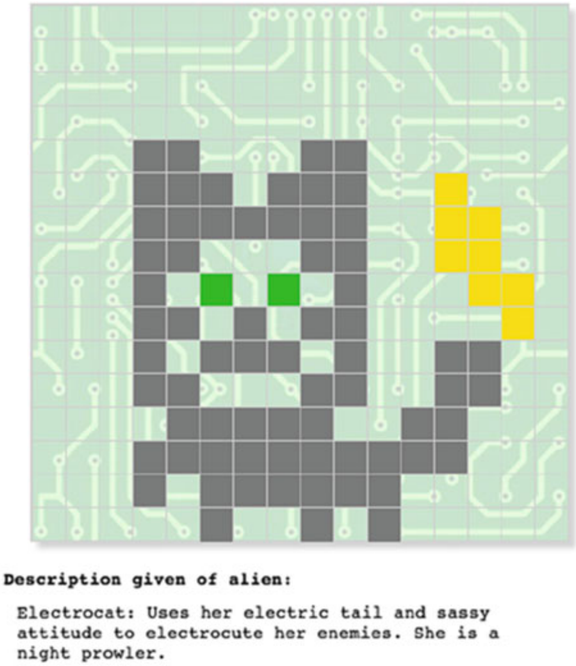
4.2 Procedure

Each participant engaged in four steps: (i) pre-survey, (ii) description of task and creature examples, (iii) 10 min task, and (iv) a post-survey. Before and after the task, participants were asked to rate their level of confidence in prototyping with electronics, and their current feeling of creativity on a 1–7 point Likert-type scale, with 7 being the highest. At the end of the study, cognitive flow was measured using the flow short scale survey developed and validated by previous studies on flow (Engeser and Rheinberg 2008). We logged usage statistics, such as the number of creatures created and the number of LEDs used per participant. For each group we computed the (Y_1) average quantity of creatures created, (Y_2) average number of LEDs used, (Y_3) average change in creative feeling between pre-task and post–task surveys, and (Y_4) average change in perceived self-efficacy.

5 Results

All participants ($N = 86$) attempted to create creatures, and the virtual circuit building exercise appeared to be an effective design task for generating diverse prototype ideas (Fig. 7). Qualitatively we found that the 16×16 grid was sufficient to express creative ideas and most participants provided colorful verbal descriptions. Statistically comparing the control group with each group (two-tailed *t*-test), we found the following results (summarized graphically in Fig. 8):

1. **Prototype Quantity:** Participants in Group B made significantly fewer creatures than the Control. (2.42 vs. 3.47, $p < 0.05$). Both Group A and B used fewer LEDs than the control group ($A = 16.6$ $B = 12.62$ vs. $Control = 27.6$, $p < 0.05$). There was no statistical difference between the prototype quantity of Group A and the Control (3.43 vs. 3.47, $p > 0.05$).
2. **Self-efficacy and Creative Feeling:** Both the Control and Group A reported an increase in creative feeling and self-efficacy after the design task (+0.57 and +0.47, and +0.5 and +0.44, respectively). Group B reported a decrease in creative feeling and self-efficacy (−0.58, −0.37, on the 7-point Likert scale), and these changes were significantly different than the Control ($p < 0.05$).
3. **Cognitive Flow:** Group A had an increased flow score compared with the Control (4.99 vs. 4.56, $p < 0.05$). There was no statistical difference between the flow scores of Group B and the Control.



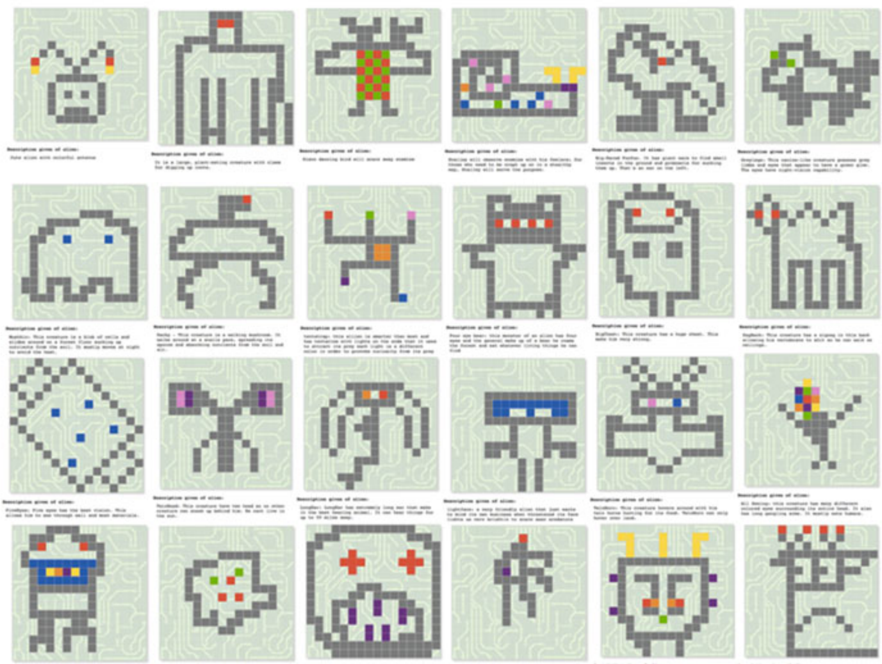


Fig. 7 A sample of creatures created by participants, illustrating the diversity of creatures

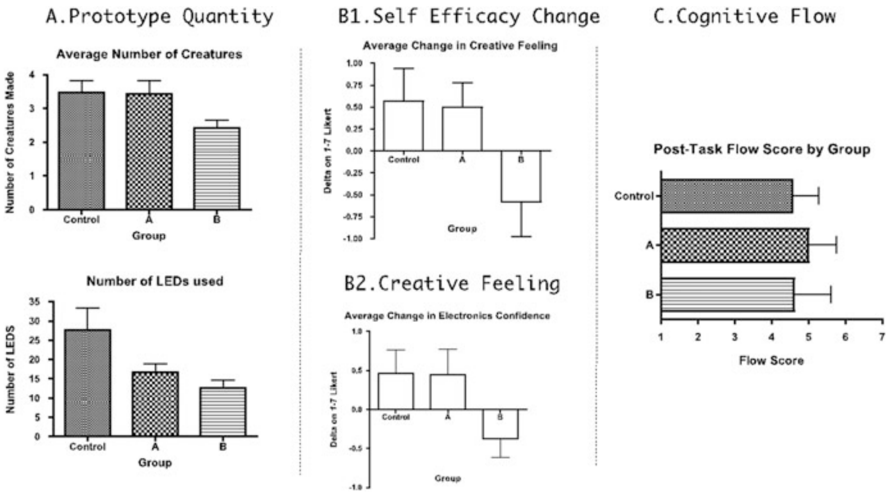


Fig. 8 Summary of prototype quantity, self-efficacy change, creative feeling and flow across groups

6 Discussion

The results of this study show that we can measurably increase a designer's perceived self-efficacy and creative feeling through the use of modular prototyping tools. In contrast, exposing designers to a higher degree of technical detail while prototyping (e.g., through having one select the correct wire orientation for each LED) had a significantly negative effect on design performance, as evidenced by a decrease in prototype quantity, self-efficacy, and creative feeling for participants in Group B. It appears that repeated exposure to technical interruptions while prototyping can create a cumulative barrier to one's design creativity. The act of wiring individual LEDs diverted time away from the task of creating, which was seen through the reduced quantity of ideas generated by Group B. However, the increase in cognitive flow observed in Group A (the group presented with a consistent wiring orientation for each LED) suggests that exposure to technical choices that require minimal cognitive interruption contributes to a state of flow, which corresponded to an increase in creative feeling and self-efficacy. The findings from this study underscore the importance of increasing the modularity of technology design tools to encourage playful prototyping and creative expression. While this pilot study shows promising results with a software simulation and crowd-sourced participants, future work will focus on matched cohort testing with physical parts, rather than simulated circuits. Also, we aim to evaluate how modularity impacts objective measures (i.e., prototype quantity, accuracy, and functionality) associated with the designs produced.

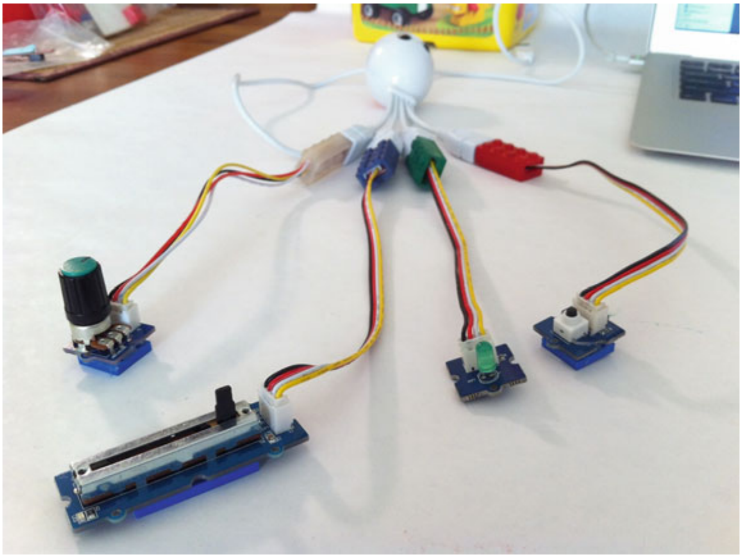


Fig. 9 The Bloctopus physical prototyping toolkit: an example of a modular electronics and software toolkit (Sadler et al. 2015a, b)

6.1 Translating to Physical Toolkits

This research highlights that modularity increases the novice's confidence to create prototypes in a constrained domain. While the prototyping task was virtual, we expect to find comparable results when adapted for physical toolkits, and this is the focus of ongoing work. The *Bloctopus* physical prototyping toolkit of Sadler et al. (2015a, b) shows an initial technical implementation of a plug-and-play electronics system that prioritizes black-box modular interactions, and interfacing with physical LEGO blocks (Fig. 9). Future work is exploring how to overcome the functionality ceiling that is linked with a black-box module approach (Sadler et al. 2015a, b). This ongoing work has highlighted that both the software and electrical domain can benefit from increased modularity. We see such *integrated* hardware-software modular interactions as the first step in a more universally accessible prototyping toolkit.

7 Conclusion

We present a creative prototyping user study with N = 86 crowd-sourced participants, on a virtual LED circuit task. The study tested the hypothesis that modules can reduce the difficulty of prototyping by hiding technical details into cognitive chunks. The findings demonstrate with empirical evidence that modularity has creative benefits on prototyping tasks. Specifically, the results show that increasing the modularity of design tools allowed participants to create more prototypes, and significantly increased designers' feelings of creativity, self-efficacy and cognitive flow. This study lays the groundwork for an understanding of how modular tools amplify creative prototyping.

References

Anderson C (2014) Makers: the new industrial revolution. Random House, New York

Baldwin CY, Clark KB (2000) Design rules: the power of modularity, vol 1. MIT Press, Cambridge

Bandura A (1977) Self-efficacy: toward a unifying theory of behavior change. Psychol Rev 84 (2):191–215

Csikszentmihalyi M, Csikszentmihalyi IS (1992) Optimal experience: psychological studies of flow in consciousness. Cambridge university press, Cambridge

Dow SP, Glassco A, Kass J, Schwarz M, Schwartz DL, Klemmer SR (2010) Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM Trans Comput Hum Interact 17(4):18

Engeser S, Rheinberg F (2008) Flow, performance and moderators of challenge-skill balance. Motiv Emot 32(3):158–172

Kulkarni C, Dow SP, Klemmer SR (2014) Early and repeated exposure to examples improves creative work. In: Plattner H, Leifer L, Meinel C (eds) Design thinking research. Springer, Heidelberg, pp 49–62

Hartmann B, Doorley S, Klemmer S (2008) Hacking, mashing, gluing: understanding opportunistic design. IEEE Perv Comput 7(3):46–54

Hartmann B, Klemmer SR, Bernstein M (2005) d. tools: integrated prototyping for physical interaction design. IEEE Pervasive Computing

Marsh RL, Landau JD, Hicks JL (1996) How examples may (and may not) constrain creativity. Mem Cogn 24(5):669–680

Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63(2):81

Pimmler TU, Eppinger SD (1994) Integration analysis of product decompositions. Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge

Sadler J, Durfee K, Shluzas L, Blikstein P (2015a). Bloctopus: a novice modular sensor system for playful prototyping. In: Proceedings of the ninth international conference on tangible, embedded, and embodied interaction. ACM, pp 347–354

Sadler J, Shluzas L, Blikstein P, Katila R (2015b) Creative chunking: modularity increases prototyping quantity, creative self-efficacy and cognitive flow. In DS79: Proceedings of the third international conference on design creativity, Indian Institute of Science, Bangalore

Tierney P, Farmer SM (2002) Creative self-efficacy: its potential antecedents and relationship to creative performance. Acad Manage J 45(6):1137–1148

Part III
Measuring Design Thinking

Measuring the Impact of Design Thinking

Jan Schmiedgen, Lea Spille, Eva Köppen, Holger Rhinow, and Christoph Meinel

Abstract This article focuses on how organizations measure the impact of design thinking. The results are based on a quantitative survey that is complemented by qualitative interviews with experienced design thinkers. Even though a majority of respondents perceive some kind of impact, only a minority has tried to determine the impact in some way. Those who do not evaluate the impact, often do not know how or lack the necessary resources. The metrics of those who do measure design thinking's impact vary considerably, but customer feedback and satisfaction is a recurring theme. We propose that the traditional means of performance measurements are often ill-suited for evaluating the impact of design thinking. We conclude with a promising industry example of how traditional measures can be used to gauge overall performance and how a story-based approach can capture the role of design thinking.

1 Introduction

When companies introduce new techniques, methods or working approaches, managers often feel the need to measure their impact. Defining parameters to measure impact is as challenging for design thinking as it is for other innovation practices (Schepurek and Dulkeith 2013). After all, innovation projects are—by definition—explorations into unknown territories with a high risk of failure (e.g., Rhinow and Meinel 2014).

How is design thinking practiced in organizations? How do people measure the impact of design thinking or if not, what are their reasons for not doing it? Do any measuring standards exist? These were some of the questions that prompted us to launch a worldwide exploratory survey on the state of design thinking in

J. Schmiedgen (✉) • L. Spille • E. Köppen • H. Rhinow • C. Meinel
HPI-Stanford Design Thinking Research Program, Hasso Plattner Institute for Software Systems Engineering, Prof-Dr-Helmert-Str. 2-3, 14482 Potsdam, Germany
e-mail: jan.schmiedgen@hpi.de; eva.koeppen@hpi.de; holger.rhinow@hpi.de; christoph.meinel@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_11

organizations in early 2014. Our quantitative approach was accompanied by several in-depths interviews with experienced design thinkers.

In this article we will focus on measuring the impact of design thinking in an organizational setting and present the most interesting insights.[1]

Organizations have plenty of key performance indicators (KPI) at their disposal in order to measure success—return of investment (ROI), net present value or customer satisfaction, to name but a few (Richard et al. 2009). In our daily work as researchers and coaches of design thinking, we often hear calls for measuring the impact of innovation-related activities. For employees in for-profit companies it often boils down to two questions:

> How can I show my manager that design thinking has an impact on the bottom line? He wants upfront proof that the 'method' works.
>
> Our current 'innovation method' gets measured along the whole (stage gate) process. How do you do that in design thinking?

Keeping in mind the wide variety of organizations with their different goals and definitions of success, we wondered what practices exist in the hope they would spark a fruitful discussion as well as inspire those who are in search for their own measurements.

The results from our survey and interviews indicate that traditional means of performance measurements are often ill-suited for evaluating the impact of design thinking. We conclude with a promising industry example of using traditional measures to gauge overall performance and a story-based approach to capture the role of design thinking in the (innovation) process.

2 Survey and Interview Sample

We chose an embedded multi-method design by combining quantitative with qualitative data (Eisenhardt 1989; Eisenhardt and Graebner 2007). The quantitative data was gathered from an extensive survey and the qualitative data was gained from semi-structured interviews with selected practitioners (half pre-selected, half from the quantitative sample).

Due to the exploratory nature of our study, the survey touches on a variety of subjects with questions both deductive and inductive in nature. The deductive (theory-based) items operationalized constructs described in the existing literature on design thinking (Carlgren et al. 2013) as well as in the organizational studies of Weick et al. (1999). The inductive (open-ended) questions were added to capture relevant aspects not yet covered by the items informed by existing theoretical concepts as mentioned above. Therefore, participants had the possibility to express their point of view in free text fields in addition to answering multiple choice questions. In order to increase the validity of the responses, we employed very few mandatory questions in

[1] The data presented in this chapter is part of a larger study on the situation of design thinking in organizations and can be accessed here: http://thisisdesignthinking.net

case they were not applicable or required information that participants did not want/ were not allowed to give. Allowing participants to skip questions entails varying numbers of respondents for each question. For better readability, we will refer to the sample as a whole as N and to all subsamples as n without further distinction. The n refer to the subsample described in the respective paragraph.

The survey was conducted with an online questionnaire designed with the research software Qualtrics.[2] Pretests were conducted using concurrent and retrospective think-aloud techniques and probing: In two iterative phases the questionnaire was tested by design thinking experts with regard to comprehensibility and coherence of the questions.

Questionnaire Sample We invited managers and employees in organizations all over the world who applied design thinking to participate. E-Mail lists, social media (such as LinkedIn groups, Twitter, facebook groups) and existing contacts were the main channels to recruit the convenience sample. A total of N = 403 people answered our call in the first half of 2014.

About half of the respondents were managers of teams or organizations (51 %, n = 235), the remainder were design thinking team members.

Most of our respondents were from *profit-oriented organizations* (65 %, n = 219,[3] see Fig. 1). Another 16 % were from *non-profit organizations*, 9 % from *governmental organizations*, and 2 % from *public-private partnerships* (PPP). Most of the 8 % that make up the *none of the above* category were universities, freelancers, consultancies and social businesses.

Organization size was recorded in terms of the number of employees (European Commission 2003). 26 % (n = 118 (see footnote 3), see Fig. 2) indicated working in a 'micro' organization with 1–9 employees, another 26 % in small organizations (10–49 employees), 12 % came from medium-sized organizations (50–250 employees), 36 % from large organizations (≥250 employees). The 36 % of large organizations comprises 23 % of organizations with 250–9999 employees, and 13 % with more than 10,000 employees.

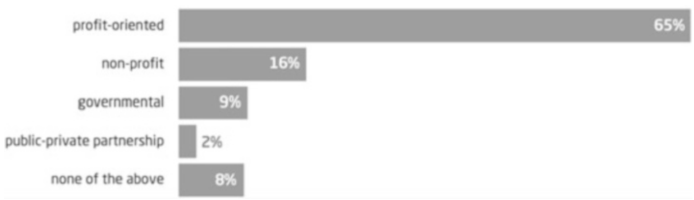


Fig. 1 Distribution in terms of corporate form (n = 219)

[2] http://www.qualtrics.com

[3] If two or more participants were from the same organization, that organization was counted once.

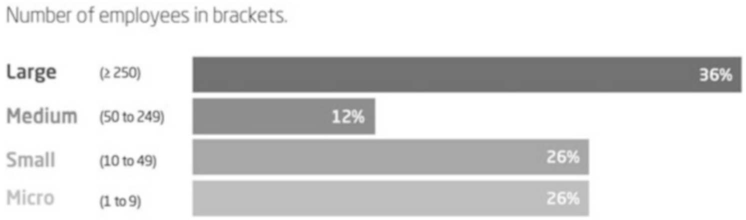


Fig. 2 Distribution in terms of organization size (n = 118)

Table 1 Distribution in terms of industry sectors (n = 147)

Information and communication	22 %
Other service activities	19 %
Education	18 %
Professional, scientific and technical activities	9 %
Financial and insurance activities	7 %
Manufacturing	5 %
Wholesale and retail trade; repair of motor vehicles and motorcycles	5 %
Public administration and defence; compulsory social security	3 %
Arts, entertainment and recreation	2 %
Human health and social work activities	2 %
Administrative and support service activities	2 %
Electricity, gas, steam and air conditioning supply	1 %
Accommodation and food service activities	1 %
Transportation and storage	1 %
Construction	1 %
Activities of extraterritorial organizations and bodies	1 %

Information and communication[4] (22 %, n = 147), *other service activities* (19 %), and *education* (18 %) were the most frequently mentioned industry sectors (UNSD 2008, see Table 1 for entire list).

If a category with an ‘other’ in the title reaches the second highest count, it certainly deserves a closer look. The group of *other service activities* (i.e., organizations offering design thinking as a service provision) included responses from “program advisory committees”, “R&D service units” and “network

[4] The strong representation of *information and communication* organizations might be due to the fact that many of our survey participants have or have had ties to the IT Systems Engineering division of Hasso Plattner Institute and therefore an immanent IT focus in their work. Another explanation could be that the IT sector faces high innovation pressure caused by digitization and user empowerment. From experience we know that design thinking heavily resonates in those (ICT) environments. Therefore it may not be surprising that the many responses from the IT sector are representative of its role at the forefront of organizations that have to react faster than others.

organizations". Additionally, there were several responses from design thinking consultancies, most of which we integrated in the more clear-cut groups when the description was unambiguous. Some respondents work in settings that might be accurately described as crossbreeds of industries, e.g., manufacturing and advisory practice ("consulting engineers") or providing half-internal, half-external support functions including design thinking services.

The responses from *education* include both educational institutions that offer programs affiliated with design thinking as well as those who do not 'sell' design thinking education and services per se. Some apply the concept to change education programs or use it to improve teaching and as a means of developing of skills and creative confidence in students.

Interview Sample In order to gain more in-depths insights, we conducted 16 qualitative interviews in eight globally acting organizations with profound design thinking experience. This sample includes a multi-sided hospitality platform, three software companies, an app developing company, a company for creative software, a hearing aid manufacturer, and a provider of electronic devices. These companies employ a variety of design thinking practitioners, most of them having gained design thinking experience for 6 years or more. One of the software companies mentioned above was Intuit, a provider of personal finance software who kindly allowed us to share their approach to measuring the impact of design thinking here.

The topics of the semi-structured interviews were based on themes and patterns that emerged during first exploration of the survey data. However, our interview guide remained relatively open to allow for the rather inductive character of our study.

If possible, all of the following survey results were gauged against the data we gained from these qualitative interviews.

3 Insight #1: Mind the Definition

Design thinking is practiced in various formats by different organizations. In order to explore the potential impact of measuring practices, one needs to put into context by whom, when, where and how something occurs that has been labeled design thinking. Our survey yielded quite a heterogeneous view on what design thinking encompasses. The most common answers (ranked in order of frequency) given by our respondents to the open-ended question asking for a definition of design thinking were as follows:

1. an iterative process,
2. a 'special' way of understanding and creatively solving so-called wicked problems,
3. user empathy,
4. a tool for collaboration,

5. a mindset,
6. a toolbox for user research and group creativity,
7. prototyping, or
8. a culture.

Analyzing the differences and nuances of design thinking definitions deserves a book chapter on its own (Lindberg et al. 2010), so we will limit this insight to mentioning the most common themes at this point. These themes illustrate that design thinking is understood and practiced in a multitude of ways in organizations. As we will see, the definitions are just as multifaceted as the approaches to measuring them. The fact that so many different definitions exist challenges the idea of a general design thinking standard of measurement.

4 Insight #2: Impact? Yes! Measure It? No

Only 24 % of our respondents (n = 167) affirmed measuring the impact of design thinking. In contrast, the percentage of those who subjectively perceive an impact is higher. 71 % (n = 181) agreed with the statement that design thinking improved their working culture and 69 % agreed with the statement that design thinking makes innovation processes more efficient. The impact seems to be most noticeable in areas that some respondents described as "intangible" or "soft facts".[5] One user described the way measurement is done as "[. . .] more intuition than statistical." Another user gave the puzzled response: "DT [design thinking] is part of the culture and the approach to work with customers. And it is adapted to our needs, approaches and mixed with other methods['] processes. So what should we measure? A mindset? A part of a method?"

There were 101 respondents who provided us with details on why they do not measure the impact of design thinking. One respondent neatly summed up the main reasons by saying "don't know how, no time and money."

What to Measure? The largest group was at a loss as far as what to measure and found it "hard to figure out the right parameters for proper management". They had "no idea," "no tool," "no KPIs available," "no suitable metric," no "test groups" or "benchmarks against which we can measure". One respondent summarized his perception by saying: "Do you measure success based on the sales success? Do you measure it based on the customer feedback. . . all those things are not really defined."

[5] For better readability we corrected obvious misspellings and adjusted initial capital/lowercase letters without comment according to the Chicago Manual of Style (2010). Additionally, we converted capital to lowercase letters within quoted passages for general terms and references to bodies or concepts, also without comment. Many of our respondents were not native speakers of English. Therefore, their statements might not always correspond to Standard English.

Too Little Experience Another group of participants was not yet able to define any measurements, because they have had “too little experience with DT so far” or it was “too early” to measure it respectively. One respondent further assumed it is a risk to bring in new parameters at an early stage: “It is still so young [new], and the innovation departments are reluctant to strangle themselves with further KPIs.”

Lack of Resources A third group of respondents wishes to measure the impact, but has a lack of resources. As one participant put it, “Our organizational unit does not pay adequate attention to KPIs, so this lack of tracking success is not specific to design thinking.” Another team tries “[. . .] to keep up with the status quo of the projects, but usually after finishing the user-centered part of it, it is unfortunately not up to us anymore”. Some also “don’t have time for it” or have “[. . .] no time and money”.

Flying Under the Radar Some respondents ‘fly under the radar’ of their management. Measuring the impact of design thinking would require a “mandate from managers”, but “design thinking is not a fully supported approach by the management in improving the working efficiency” or “DT has never been introduced formally.”

Working Culture For those who consider it part of their working culture, one respondent stated that the “benefits of design thinking in our organization are in most cases intangible” and another respondent wondered: “How would you measure enhanced teamwork?” Design thinking was also found hard to measure since there are “too many soft facts involved.” Furthermore, one said “it is too much about feelings, soft-skills and culture and hard to measure.”

Unnecessary or Impossible to Measure Then again, not everybody is convinced of the necessity to measure the impact of design thinking at all, because it is just “not needed yet,” or they have “not thought about it yet.” One person simply said: “Why should we?” while another is “not a fan of measuring.” Measuring design thinking’s impact is also of little importance when it “is not viewed as a ‘real’ process”, but rather considered to be “[. . .] an unproven theory.” Others consider it impossible to measure, because it “is a mindset[,] either you have it or not” or “it’s one method between others.”

5 Insight #3: Not All Consider Their Measurements to Be Valid

Surprisingly, quite a few people first negated measuring the impact of design thinking. But when asked “why not?” they mentioned actions that indicated measurements of all sorts, even if rudimentary. After answering ‘no’ to the question: “Do you measure the success of design thinking in your organization?”, one respondent tacks on a measurement almost as an afterthought: “We are the first company in our market applying design thinking. So first we encourage and educate the market. We measure the results and feedbacks from design thinking related events organized by us [. . .]".

This respondent is not alone, as the following examples illustrate:

"Why *don't* [emphasis added] you measure the success [of design thinking]?"

- "How? Of course, we get feedback from our clients, and track their development, but other than that?"
- "Because our department is measuring it by business cases that are generated by it, but not the application of it by itself."
- "We don't know how to measure it in terms of KPIs. We collect feedback, on whether we BELIEVE it worth-while investing all the love and resources that we invest, and then adapt accordingly [. . .]."
- "It does not really make sense, but we collect good examples and cases to show to others."

The mindset of these responses suggests a rather paradoxical situation. It seems that some teams have found measurements, in the broadest sense, they deem useful (since they are in use), yet found themselves unable to elevate these same tools to the status of "measurement-hood." At the same time they are lacking in what they might consider a valid measurement. Do they not trust their own judgment? It seems challenging to people to define factors that are important for assessing the effect of design thinking.

6 Insight #4: Design Thinking Is Measured in Manifold Ways

Of those 40 participants who indicated that they do measure the impact of design thinking, 23 participants provided details into the 'how'. A few answers hinted at operationalizations (such as "stunning ideas per hour") and many seemed to rely on reflection.[6]

Customer Feedback One theme that recurrently emerged was customer feedback and satisfaction. Knowledge was gathered "internally: customer centricity score [. . .]; externally: net promoter score, KPIs of units, customer satisfaction [. . .]" or with "client feedback" as well as "responses to campaigns" and "[. . .] brand value perception from the user's perspective." Others focused on "usability metrics, how [design thinking] affects the user experience" and "usability measurements in terms of efficiency, effectiveness and satisfaction, customer validation[,] interview results, number of live customers." One respondent even commented: "In my

[6] Unfortunately, some measurements were described in terms so vague that they were impossible to interpret.

experience KPIs are meant to deliver exactly the message the measuring party wants to deliver. I prefer customer/user feedback and testimonials." These quotes show that the user feedback is a common evaluation tool for design thinking. Consistent with research on empathy with the user (Köppen and Meinel 2015), Wendy Castleman from Intuit emphasized the importance of the user as well. She said: "I would say the [companies] who have any success are focused on their customers. I do talk to companies frequently. They are interested in the innovations. They want things that are going to be successful and bring them to a new level. But if they don't actually care about their customer experience it isn't [going to] work. Because in order to really be revolutionary and different and inspirational in things that people buy, it's [got to] solve their problem, right?"

Design Thinking Activities Some respondents measured 'design thinking activities' such as "[number] of projects, trainings, trained people, junior coaches", "how many employees are formally trained", or the "number of coaches, number of projects".

'Immediate' Results Some participants went one step further by measuring the 'immediate' results of their design thinking projects, manifested in: the "number of transferred innovation to development," number of "concepts finish[ed] in actual strategic plans," and "outcome of DT projects, e.g., launch of application." One participant mentioned the "number of projects that get second round funding", yet adds: "It's a bit manufactured [fabricated], but since we need a KPI. . ."

Traditional KPIs Those who took a closer look on more traditional KPIs focused on "financial performance," "market success," "success and revenue of projects initiated by DT," and "the ROI on each project undertaken," and "increased sales."

Reflective Measurements Questionnaires and other reflective measurements are not uncommon, taking the form of "evaluation forms", "survey[s] of depth of knowledge and application of process and tools in the organisation, qualitative feedback during and after projects" or "questionnaires after each DT process step."

Working Culture A few directed their attention to the impact on working culture and measured "motivation, effectiveness, team collaboration [and] engagement".

7 Insight #5: The Impact of Design Thinking as a Butterfly Effect

When so many of our survey respondents seemed at a loss as how to measure the impact of design thinking, they touched upon an important topic. The problem with creating metrics for design thinking in the context of a larger innovation effort can be described as a so-called butterfly effect.[7] With so many factors influencing the success of any given project—from different objectives or a different team composition to the world's economic situation—how can one hope to single out one definitive outcome caused by design thinking?

One survey participant summarized the dilemma eloquently: "How would you measure it? Measuring implies that hard metrics can be derived and tested against competing methodologies. Besides, if design thinking is embedded in your organization it cannot be measured as a single concept. Therefore we measure our general performance with several KPIs, but we cannot determine specifically to which level design thinking contributed to this."

All responses considered, our conclusion at this point is that there is no specific stand-alone design thinking KPI. Furthermore, it is highly probable that there can be none capable of adequately determining the specific impact or contribution of design thinking in an organizational setting due to the complexity of the situation.

However, difficulties in finding hard metrics is not the end of the story. Intuit is one of the experienced companies with regard to design thinking in our interview sample. Their approach to measuring the impact of design thinking is remarkable, as it tackles many of the doubts and challenges mentioned by our respondents.

8 Insight #6: Story-Based Approaches: A Silver Lining

When asked how the impact of design thinking was measured, two participants responded by "success stories" and "we collect good examples and cases to show to others". What does this mean?

We spoke with Kaaren Hanson, Vice President of Design at the time, and Wendy Castleman, head of the 'catalyst program'[8] from Intuit, a software provider of business and financial management solutions for small and mid-sized businesses. They provided us with details on their version of a story-based approach. They internally showcase success stories in books produced each fiscal year. Each story enables the reader to trace back case-specific measurements, which often emerge alongside the individual project.

For Intuit, isolating and quantifying the specific impact of design thinking has turned out to be impossible because "it has just permeated what we do and how we do it now. So, [... design thinking] becomes such a part of the company [that] you can't separate it out and measure it per se." They went so far as to conclude that "the

[7] The *butterfly effect* is a concept stemming from chaos theory. It postulates that small events in a complex system can have large, widespread consequences. In hindsight, it is nearly impossible to isolate specific causes (Hilborn 2004).

[8] The catalyst program allows employees with regular jobs but trained in design thinking—so-called innovation catalysts—to spend 10 % of their working hours to coach workshops.

way we look [at] success is not so much about metrics. There are so many things that can confound a metric."

Even though they accepted that quantifying the impact was a futile endeavor, they still strived to capture the role design thinking played in their work. So instead, they chose to take a qualitative approach. Now they create stories that can be shared within the organization and allow other departments the opportunity to see and understand the impact of design thinking more easily and share the success on a different level: "[. . .] I think stories are in fact the most impactful [. . .]. Anecdotes are powerful."

Yet the cases are more than a mere collection of anecdotes. Every single one contains the important measurements the company cares about like revenue, cost, profit, employee engagement, and customer engagement. But measurements need a context for people to relate to—so they wrap them into stories.

Before developing their story-based approach, Intuit went through several stages of measuring that sounded very similar to those our survey respondents described. They started off by measuring activities ("How many people are doing this?") and examining several outcomes like increased revenue, decreased cost, increased employee engagement, or increased customer experience. Wendy suggests: "First measure activity, then start to measure impact with whatever your company cares about." But the types of measurement they start with do not always remain the same.

For Intuit, design thinking has a profound impact on the choice of appropriate types of measurement. When exploring the problem at hand ("problem reframing"), initial metrics can become obsolete. One key to Intuit's successful practice is allowing new measurements to emerge during the process. Kaaren experienced that "often people want a measure[ment] from you that you can't provide them with. But you can give them something else." She remembered one case where "the metric that mattered to the team when they started and what they were asking for was 'increase conversion'. The metric they ended up getting (e.g., customer and employee engagement) was something completely different. But everything felt much better about it." Additionally, allowing new measurements to emerge can also lead back to initial metrics, because in the example above "ultimately [. . .] it increased conversion!"

Apart from allowing measurements to emerge during a project, Wendy offered another piece of advice: "People ask about our innovation program and design thinking all the time. Often they are just interested in the innovations. They want things that are going to be successful, which brings them to a new level [. . .]. But if they don't actually care about their customer experience it isn't [going to] work. [. . . You] have to understand what your company culture values and make sure you align to that first!"

The way Intuit solved their problem struck us as remarkable because it contained elements from all the other insights. It describes the difficulties of quantifying the impact of design thinking, especially the more it is integrated into the entire organizational culture. Over the years Intuit tried various ways to measure the impact of design thinking. Finally, they reached the conclusion that isolating a quantifiable impact was impossible. With their story-based approach they found a balance between quantitative measurements for overall projects and qualitative ones to describe the role of design thinking in the innovation process.

9 Conclusion

In this contribution we explored if and how organizations measure the impact of design thinking. Our data yielded five main insights:

1. First of all, many very different practices are labeled design thinking—making them challenging to analyze.
2. Even though many respondents report some kind of impact, very few actually measure it.
3. Some utilize evaluative tools but do not seem to consider their tools to be valid (or to be 'real' measurements).
4. Those who do measure the impact of design thinking have manifold of ways of doing so, even though some of the methods seemed 'a bit manufactured [fabricated]' as one respondent commented. The strongest measuring theme was customer feedback and satisfaction.
5. The impact of design thinking is very difficult to quantify and appears to be a so-called butterfly effect.
6. A story-based approach can mitigate the difficulties of trying to single out the role design thinking plays in a project.

Measuring metrics are important for many and challenging for all organizations. Furthermore, the frequently cited lack of ideas for KPIs indicate that traditional means often fall short of assessing the impact of design thinking. Companies that nevertheless want to evaluate the effect have to be flexible and adapt.

Finding measurements for design thinking resembles the process itself—it is iterative.

The following suggestions are an appeal to common sense and an encouragement to reframe problems, explore the problem spaces and integrate measurements and metrics that justifiably present themselves after reframing the problem at hand:

- **Start with what you've got:** When beginning to evaluate the impact of design thinking, start with the metrics your organization cares about and/or start by measuring design thinking activities (e.g., number of projects or coaches in order to record what is happening).
- **Allow measurements to emerge:** Intuit learned to embrace detours. Limiting your problem space to the initial objective also limits your chances of finding related measurements that might serve as a stepping stone to reach your initial goal.
- **Evaluate metrics pertaining to the 'bigger picture':** Measurements must be observed within their context. Keep questioning whether your current

measurements are still valid. Our interview partners at Intuit prompted the following example: One of their KPIs is the *net promoter score* (NPS) to measure client satisfaction. The score represents the ratio between so-called promoters (who indicate that they are very likely to recommend your company/brand/product X to a friend/colleague/relative) and so-called demoters (who indicate that they are not likely to make a recommendation). Picture the following situation: An organization just made major improvements to their product and raised the price when a financial crisis hit. What would be likely to happen? Customers complain that the product is too expensive, so they are less likely to recommend it to someone else. As a consequence, the NPS goes down—but does that mean they should not have improved their product? Most likely, the NPS gets a dent for a while but improves in the long run—when people start appreciating the improvements. So unless a metric is extremely robust, relying on a measurement without considering the bigger picture is risky, sometimes even harmful.

Most of our respondents were from for-profit companies, so many of our examples are too. However, hopefully for-profits and non-profits alike (and all of those in between) will find useful information on measuring the impact of design thinking in an organizational setting.

10 Limitations of the Study and a Note on Future Research

We approached the subject of measuring the impact of design thinking by asking open-ended questions and assigning *a posteriori* categories. While this is a reasonable course of action for an exploratory study, it bears the risk of post hoc rationalization (i.e., after the fact justification). Future research is necessary to test whether our results can be reproduced.

Furthermore, it seems that not everybody qualifies their acts of measuring as such. Thus, they negate the direct question of whether they measure the impact of design thinking. Instead of relying on a common understanding of an abstract term like 'measurement,' questions describing concrete actions[9] can shed light on this area. A study using action-based questions would possibly find a higher rate of people who measure the impact of design thinking.

Acknowledgments All results are part of a survey on the impact of Design Thinking in organizations. We would like to thank Prof. Hasso Plattner and the HPI-Stanford Design Thinking Research Program Committee for their support. We thank Dr. Sharon Nemeth for her editorial support.

[9] Questions based on concrete action could be: "Do you record the number of design thinking projects?" or "Do you collect customer feedback?"

References

Carlgren L, Elmquist M, Rauth I (2013) Demystifying design thinking: a conceptual framing of design thinking in use. Draft, Goteburg

Eisenhardt K (1989) Building theories from case study research. Acad Manage Rev 14(4):532–550

Eisenhardt K, Graebner M (2007) Theory building from cases: opportunities and challenges. Acad Manage J 50(1):25–32

European Commission (2003) Recommendation 2003/361/EC: SME definition. http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/sme-definition/index_en.htm. Accessed 18 Nov 2014

Hilborn R (2004) Sea gulls, butterflies, and grasshoppers: a brief history of the butterfly effect in nonlinear dynamics. Am J Phys 72:425

Köppen E, Meinel C (2015) Empathy via design thinking: creation of sense and knowledge. In: Plattner H et al (eds) Design thinking research. Springer, Heidelberg

Lindberg T, Noweski C, Meinel C (2010) Evolving discourses on design thinking: how design cognition inspires meta-disciplinary creative collaboration. Technoetic Arts 8(1):31–37

Rhinow H, Meinel C (2014) Design thinking: Expectations for a management perspective. In: Plattner H et al (eds) Design thinking research: understanding innovation. Springer International Publishing, Heidelberg, pp 239–252

Richard PJ, Devinney TM, Yip GS, Johnson G (2009) Measuring organizational performance: towards methodological best practice. J Manage 35(3):718–804. doi:10.1177/0149206308330560

Schepurek S, Dulkeith E (2013) Innovation performance measurement: KPIs for goal-setting. Paper presented at the XXIV ISPIM conference—innovating in global markets: challenges for sustainable growth, Helsinki, Finland, 16–19 June 2013

United Nations Statistics Division (2008) International standard industrial classification of all economic activities (ISIC), Rev.4 classification scheme for determining the industry sectors. http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=27&Lg=1. Accessed 18 Nov 2014

University of Chicago (2010) The Chicago manual of style, 16th edn. University of Chicago Press, Chicago

Weick KE, Sutcliffe KM, Obstfeld D (1999) Organizing for high reliability: processes of collective mindfulness. Res Organ Behav 1:81–123

Developing Design Thinking Metrics as a Driver of Creative Innovation

Adam Royalty and Bernard Roth

Abstract The creative behaviors that underpin design thinking are difficult to measure. This is problematic because people who have a desire to practice design thinking in an organizational context are often assessed only on their ability to execute via traditional metrics. Therefore they have less incentive to work in a creative way. In order for organizations to fully support and incentivize design thinking, they must measure creative behaviors as much as they do executional behaviors. This chapter highlights a suite of initial metrics that arose from research on d.school alumni and organizations applying design thinking as a core driver of their innovation strategy.

1 Introduction

What Gets Measured Gets Done This popular business adage reveals a barrier many people face when applying design thinking. Most organizations rely on efficiency and productivity to succeed. As such, measurement is geared towards identifying and rewarding employee behaviors that support these execution-oriented goals. Most organizations that turn to design thinking as a methodology for innovation know the process will be more exploratory and less focused on execution. This means that they need to identity and reward different employee behaviors—more creative behaviors—that support creativity-oriented goals. However, the vast majority of organizations continue to utilize execution-oriented measures without adding any creativity-oriented measures. This results in two major problems. The first is that most organizations lean on metrics, at least in part, to determine the success of new initiatives. If design thinking cannot demonstrate success in a measurable way, it may be abandoned for other innovation methodologies that are easier to track (Nussbaum 2011). The second major problem

A. Royalty (✉) • B. Roth
Hasso Plattner Institute of Design (d.school), Building 550, 416 Escondido Mall, Stanford, CA 94305-3086, USA
e-mail: aroyalty@stanford.edu

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_12

is that employing only execution-oriented measures disincentivizes the application of design thinking because only behaviors leading to better execution will be rewarded. Previous work has shown that many of these measures simply do not exist yet (Royalty et al. 2015). Developing a set of usable creativity-oriented measures that capture how well individuals and teams learn and apply design thinking should address these problems.

This chapter describes two lines of research aimed at developing useful creativity-oriented measures. The first study further explores the creative agency scale (Royalty et al. 2014) by using a robust control group. The goal is to demonstrate the internal impact of design thinking for people who take a Stanford d.school course. The second study highlights the development and testing of creativity-oriented metrics that could be used within organizations to measure and promote the application of design thinking. Both of these objectives fit a common theme of understanding the difference between design thinking and other problem solving processes and mindsets. Furthermore, these metrics move us one step closer to answering an even larger question: does design thinking have a measureable advantage over other working processes? If this can be convincingly shown, the design thinking movement will have responded to a major critique. This should make it easier for d.school alumni and other design thinking practitioners to communicate the value of this way of working to their colleagues and organizations.

2 Study 1: Creative Agency

2.1 Background

One of the fundamental goals of the d.school is to instill *creative confidence* in its students (Kelley and Kelley 2013). But what does this really mean? Can we define it in more precise terms? Because this is regarded as an internal construct, it makes sense to turn to the field of psychology for guidance.

Creative confidence is often linked to the notion of *self-efficacy* developed by Albert Bandura (Bandura 1977). He defines it as people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives (Bandura 1994). Most psychologists believe that confidence is a different construct than self-efficacy (Schunk 1991). Confidence is a general belief in one's self, whereas, self-efficacy is more subject matter specific. For example, someone may have high self-efficacy in music but low self-efficacy in math. That same person may also have a high or low general self-confidence. This begs the question, is creativity a subject or something larger?

Even if we accept creativity as a subject, and use the term creative self-efficacy, it still is not clear that creative confidence is the same thing as creative self-efficacy. Self-efficacy might be an important aspect but not completely equivalent. Another goal of the d.school is to help students apply design thinking in real world contexts. In fact, that is the basis behind the executive education program. Given this, creative self-efficacy seems to provide an incomplete definition. After all, if one has all the belief in the world about their creativity but never applies it, the d.school probably would not say that person has a lot of creative confidence. That is why we developed *creative agency*, the ability to apply one's creativity, as a complement to creative self-efficacy (Royalty et al. 2012). We believe the two together make up creative confidence.

Creative agency was developed and turned into a scale in previous HPDTRP projects (Royalty et al. 2014). It consistently demonstrated a significant change in people's dispositions before and after design thinking learning interventions. However, we never tested it with a sufficiently large control group. This year we identified two separate control groups for a quantitative, survey-based study. The treatment group was comprised of students who applied and enrolled in the 10-week introduction to design thinking "Bootcamp" course taught at the d.school during the fall quarter. One control group we identified was the set of all students who applied to the Bootcamp course but were not admitted. In most cases, the students not accepted were not necessarily less qualified; there simply were not enough available spots in the course. The other control group was made up of students who took a more traditional product design course on innovation and design. This course is also in the school of engineering and helps students generate new solutions to address an unmet need. The instructors, however, teach a different process than the d.school. The course is taught every quarter of the academic year, but not every student takes all three quarters. We only surveyed students from the fall quarter because that matched up with the fall quarter Bootcamp course and because the final two quarters have more of an implementation focus, which extends beyond the scope of most d.school courses.

In addition to capturing creative agency, we also surveyed for *creative growth mindset*. This is an adaptation of Carol Dweck's work on mindsets (Dweck 2006). She has shown that people who believe that intelligence is malleable (growth mindset) tend to perform better than people who believe that intelligence is static (fixed mindset). The survey used to detect a growth mindset is well established and has been modified to measure growth mindsets in other areas. However, there is no research yet using a modification of this survey to measure a creative growth mindset.

This construct is interesting for at least two reasons. One is that it can potentially shed some light on the people who choose to apply for d.school courses. One might expect people who applied for the Bootcamp course to have a more malleable view of creativity then those who took the Smart Products course and did not apply to the Bootcamp course. In other words, they believe they can enhance their creativity and that is a reason why they take a d.school course. The second interesting point is that a creative growth mindset may initially not be very high in d.school course applicants but may increase during the course of their experience. In this case, we would expect to see a change in the pre and posttest for students who took Bootcamp.

It is important to note that the psychological constructs listed above may very well affect how people continue to learn and apply design thinking. Self-efficacy has been shown to lead to increased motivation and persistence (Schunk 1991). These traits are especially important for students who graduate and struggle to use design thinking in a real world context. In most cases, organizations have not learned how to support design thinking. If we boost graduates' creative self-efficacy, they should be more resilient as they attempt to practice this methodology in companies that are not as nurturing an environment as the d.school. This is important because the success of design thinking relies on people applying it beyond the d.school.

Similarly, people with growth mindsets respond better to failure (Dweck 2000). They see it as an opportunity to improve rather than a judgment of their character. That suggests that the higher the creative growth mindset, the more likely alumni are to continue to leverage their creativity in the face of failure.

2.2 Materials

We created both an online and paper based survey. The creative agency scale is comprised of 11 five-point Likert items. There were 3 six-point Likert items on the creative growth mindset scale. The remainder of the survey featured demographic questions.

2.3 Procedure

A link to the online version of the survey was added to the end of the Bootcamp course's web-based application. Over 120 applicants completed the survey. The Bootcamp teaching team admitted 33 students who completed the survey. On the final day of class (the final day of the fall quarter), the students were given the paper-based version of the survey. This yielded 31 Bootcamp students who took both the pre and the post survey. The online version was emailed to all the applicants who took the pre survey but were not admitted into the course. A total of 51 subjects responded. None of them enrolled in any d.school course during the fall quarter. This is not surprising because there are very few other d.school courses taught in the fall.

The paper-based version of the survey was given during the first and last weeks of the fall quarter of a product design course. This course covers topics related to innovation but does so using a more traditional educational model. From that course 31 students out of approximately 45 enrolled completed both the pre and post survey.

At the end of the 2013–2014 academic year, an online version of the survey was sent to subjects in all three conditions who completed both the pre and post survey.

Table 1 Pre and post survey averages

	Bootcamp applicants (enrolled)	Bootcamp applicants (not enrolled)	Smart products (enrolled)
Pre survey	Moderately confident	Moderately confident	Moderately confident
Post survey	Very confident	Moderately confident	Moderately confident

The creative agency scale captures people's confidence in applying their own creativity. The average pre and post survey response of subjects in the three different conditions is listed in Table 1.

The results of the end-of-year survey are still under analysis. One of the major questions is if creative agency is sustained months after completion of a d.school course. This would fit with previous results (Hawthorne et al. 2014). However, it is not yet know how many respondents in any of the conditions enrolled in d.school courses during the winter and/or spring quarters. It may be the case that we will also have enough data to explore the effect taking multiple courses has on creative agency changes.

The creative growth mindset measure computes a score between 3 and 15. There was little change between pre and post survey averages for the three conditions. Additionally, there was virtually no difference in the averages between students who took Bootcamp and those who did not. However, students who applied to Bootcamp average a point higher than those who took the traditional product design course; 13–12.

2.4 Discussion

The survey analysis suggests that taking Bootcamp did have a significant effect on students' creative agency. It is also interesting to note that none of the pretest averages were significantly different from one another. This indicates that creative agency is not a predictor of who is going to apply for a d.school course.

It appears that creative growth mindset does not significantly change in any condition. Though this might be due to a ceiling affect as most students exhibited moderate or high growth mindset. Still the applicants to the Bootcamp course did have a higher creative growth mindset. This is not completely surprising as one would assume that students taking a d.school course would be doing so in part to work on their own creative capacity. Still having a high creative growth mindset in of itself does not lead to an increase in creative agency. It might be the case that a high creative growth mindset is a necessary but not sufficient requirement for boosting creative agency. A next step would be to investigate the creative agency changes in Bootcamp students with low initial creative growth mindset.

3 Section 2: Design Thinking Measures in Organizations

3.1 Background

If instilling a sense of creative confidence in students is the fundamental goal of the d.school, then the impact of the d.school should be measured by how well alumni apply their creativity to solve real world problems. If that is the case, then we need design thinking measures that can capture how this methodology is applied in organizational contexts. To do this we focused on teams as the unit to measure because many organizations envision teams using design thinking over the course of a project as a core component of their innovation strategy.

Previous HPDTRP work has shown that design thinking is contextually dependent (Marelaro et al. 2015). This suggests that design thinking measures are also contextually dependent—there are no perfect universal measures. Fortunately there are three principles to abide by when creating design thinking measurement tools for organizations (Royalty et al. 2015): (1) they must be easy to use by employees, (2) they must be calibrated to the organization's goals for using design thinking, and (3) internal design thinking experts should, when possible, oversee the use of these measures in order to verify that they are being used accurately.

Guided by these principles, we used a design-based research methodology (Barab and Squire 2004; Glaser & Strauss 2009) to engage with four organizations to develop design thinking measures. Each corporate partner has made a commitment to design thinking. We classified a commitment as having an official company-wide mandate to use design thinking as part of their innovation process (i.e., Courage 2013). Furthermore, each organization must have employees trained in design thinking. The organizations we worked with are: A large IT firm, a large retail firm, a large financial services firm, and a large transportation firm. It is important to note that these companies all sit in different sectors. They all have different business models and different end users; some focus on B2B (business to business) while others are B2C (business to consumer) companies. This shows the variety of organizations interested in design thinking. All four have either hired Stanford d.school alumni or sent employees to the d.school executive education program.

Each of these organizations has a group responsible for driving internal innovation. One company calls them *innovation catalysts*. For simplicity's sake, we will use that term to refer to the innovation group members in each company. The innovation catalysts are the ones spreading and supporting design thinking as part of their innovation work. They are the primary point of contact with this research, which makes sense, because they understand design and they have incentive to use design thinking measures to assess the teams they work with. They are also the ones we co-designed the measures with.

Coincidentally, as the year unfolded these four organizations created a collaboration with the initial purpose of hosting joint training sessions. In the first session, innovation catalysts from all four companies served as design thinking coaches while training 10–12 associates from each company in mixed teams. This session was hosted by the financial firm in April. The retail firm will host the next session. An emergent goal is to find a large problem that the four-company collaboration can work on together using design thinking. This is in line with one of the original promises of design thinking to solve large, complex problems (Cross 1990). Although this collaboration did not change our fundamental research question, it did provide an additional opportunity as we were asked to help them measure the impact of this collaboration in addition to the impact of design thinking in each organization.

3.2 *Procedure*

We began this project by addressing our second design principle; understanding the goals each company has for using design thinking. Although each organization is a different context for design, we decided that because they have enough in common to work together, there might be patterns across the organizational goals that we could leverage to design measurement tools for all four to use. Two different interventions led to our understanding of these organization goals.

The first intervention was eight semi-structured interviews of innovation catalysts. We asked them why their companies began using design thinking and how that has evolved over time. Essentially we asked them about the goals of the organization. The second came from 17 reflection sheets designed to uncover how innovation catalysts teach design thinking (see Fig. 1). Each sheet is a blank timeline of a given employee's journey through a design thinking training. The innovation catalysts filled in the journey of learning design thinking they

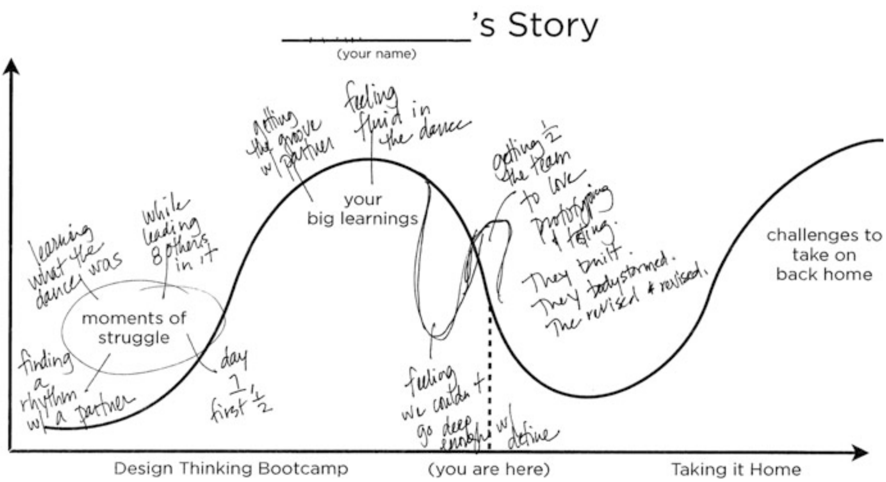


Fig. 1 An innovation catalyst's training reflection sheet

attempt to take their colleagues through. The idea is that the goals of the design thinking trainings should align with, or support, the design thinking goals of the organization.

Using these data, the initial analysis of which is included in the following section, we discovered general categories to measure. From there we worked with the innovation catalysts to modify existing measures and try new ones that we mutually felt were beneficial to the companies' innovation goals.

Our aim was to develop measures that are quantitative in nature and output to an ordered variable. This is important because we want the ability to rate teams as high, medium, or low. This is not to say that there should be no qualitative component. In fact we believe that qualitative measures could—and should—be added, just not relied on exclusively. Finally, these measure are meant to be easily captured throughout the course of a single project. Once a new project starts, the measures reset.

One important thing to note is that these measures are not necessarily meant for all teams in an organization, just the teams slated to use design thinking. There are a number of projects where teams implement traditional methodologies and successfully achieve their non-design thinking related goals.

3.3 Results

Three large themes around why these companies are turning to design thinking immediately immerged:

1. A company wide disconnect from the end user.
2. Fear of a startup taking future business.
3. Desire for teams to work in a more innovative way.

Each organization feels that in one way or another, their project teams do not understand what their customers want or need. Here is an example from an innovation catalyst: "people don't trust financial institutions after 2008. Even though [we aren't] the one to blame, they don't trust us. We need to understand what is behind that distrust in order to better relate to our customers."

Fear is an unexpectedly powerful motivator for design thinking. Two of the most senior innovation catalysts shared that their organization is afraid of newer, more nimble companies cornering the market by discovering the next big breakthroughs and taking all of their business. They believe that entrepreneurs are more likely to identify opportunities and successfully take risks.

Finally there is a desire for teams to collaborate in a more nibble and entrepreneurial way. Numerous reflection sheets indicated that one of the training goals is to teach employees the ability to flare and focus as part of the working process. Apparently focusing is much more common than flaring.

These three insights inspired four categories of measures. The categories are linked to design thinking principles that the four organizations adhere to. Below is a list of the measures with a brief description.

3.3.1 Empathy Measures

The degree to which teams are connected to their end users can be captured by the amount of meaningful contact they have with their customers. The following measures aim to capture how close teams are to users.

Measure 1E The number of days gone without interacting with a customer. An innovation catalysts has teams keep a running total of the number of days they have gone without interacting with a customer (either conducting an interview, an observation, or testing a prototype). As soon as they interact with a user, they can reset that number to zero. This measure captures the duration the team goes without user input. Alternatively, it would be possible to capture the percent of days teams interact with a customer, but we feel the former is stronger because it provides stronger formative feedback.

Measure 2E The number of users spoken with. The team keeps a running list of the people they interview or test a prototype with over the course of the project. This simply shows the amount of empathy being done. If one looks at empathy as a source of ideas, then this translates into fluency (Runco 1999). The distribution of user engagement could also be interesting. A team might start with a lot of different empathy subjects in the beginning and then hone down to few users at the end. That is why it is important to capture engagements during the entire duration of a project. This measure is a case where qualitative data, namely what was learned from each user, could be very beneficial to the team (and researchers).

Measure 3E The number of categories of people spoken with. The team creates a list of persona types that they interact with during the project. The teams generate the categories themselves. An example of such categories might be; single mothers, elderly couples, unemployed millennials, etc. This is valuable because it indicates how diverse their human centered work is. Connecting with customers means connecting with both existing *and* future customers, something this measure promotes. Again looking at empathy as a type of idea generation, this could be seen as flexibility (Guilford 1968). As with measure 2E, capturing qualitative data about the learnings from each category is highly encouraged.

3.3.2 Reframing

The ability to identify new and valuable opportunity spaces is a major component of innovation. Startups do this well in part because they are nimble enough to constantly change the problem they are working on.

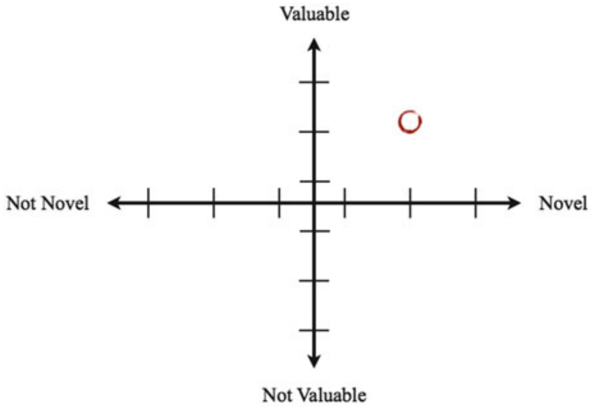


Fig. 2 Novelty/Value grid

Measure 1R The novelty/value grid is a tool that classifies the project objective. Because the objective is expected to evolve as the project advances, this tool should be used at regular intervals. The grid is comprised of two axes. The horizontal axis is how novel this objective is. Has a project like this been worked on before? Does it take the company into a new space? This doesn't necessarily mean the creation of a new good or service. It could also represent repurposing a solution in a novel context (Amabile 1996). The key feature is that it extends the company into a new space. The vertical axis is perceived value. How valuable is this objective to the company? Individual team members plot the project objective somewhere on the grid (see Fig. 2). This generates both a novelty component and value component score from each team member. An aggregate score for each axis can be calculated by taking an average or a weighted average across all respondents. Employees may feel uncomfortable rating the value of their project, so this measure is intended to be filled out anonymously.

3.3.3 Iteration

The desire for teams to work in innovative ways led to two categories. The first is iteration, or how robust their prototyping process is.

Measure 1I The number of prototype iterations. This measure captures the number of iterations performed by each individual or small group. All four organizations are pushing for more prototyping. In general, frequent iterating has been shown to lead to stronger prototypes (Dow and Klemmer 2011). Depending on the project, it might make more sense to capture the number of prototypes per feature. For example, one team may work on a small app while another may work on an entire website. The latter team then has more opportunity to iterate because there are more features. In that case it would make more sense to calculate iterations per feature. To capture this, team members simply list each iteration they create and what they hope to learn from it.

Measure 2I Number of prototypes worked on in parallel. Working on parallel prototypes leads to stronger outcomes compared to working on prototypes in series (Dow et al. 2010). As teams capture their iterations, they list them as "open" or "closed." Open means they are still actively working on that particular prototype. Closed means they are finished working on the prototype. The number of open prototypes at a given time yields the parallel prototyping score.

3.3.4 Team Collaboration

The final measure captures how well teams are working using design thinking. For this measure we turn to the Interaction Dynamics Notation tool created by Neeraj Sonalkar and Ade Mabogunje. Although this work is still developing, it would be possible for the teams to capture video of themselves and send it in for analysis.

Each measure, with the exception of the Interaction Dynamics Notation, has been developed in tandem with the innovation catalysts. Although no single measure is perfect, we believe that a suite of measures can identify the difference between teams using a strong design thinking process and teams using a week design thinking process.

Because these companies are located across the United States, it is not feasible for a research team to administer the measures. Therefore, the innovation catalysts will ultimately be responsible for distributing the measures and collecting the results. Fortunately they are motivated to participate since showing the impact of design thinking is a major part of their job. A part many have struggled with up until now.

Before the initial trial teams are selected, we will interview the innovation catalysis and ask them to map out the design thinking ecosystem in their organization. Besides providing interesting structural information about how design thinking fits in their organizations, it will help us locate teams to test with. Our goal is to prototype the measures on teams that have both a high and low amount of comfort with design thinking.

3.4 Discussion

We hope to answer a few major questions with this trial. The first is if the tools are easy and short enough for teams to complete. The second is whether or not the formative evaluation of teams allows the innovation catalysts to provide better support. The final question is what form factor makes these measures most effective. Should they be digital or paper based? Are they part of a large dashboard that fits on a wall above a whiteboard? We should have some insight into these questions after the first tests. Once these tools are fine-tuned, the next step is to compare teams' process to their project outcomes.

This work is important because the very act of measuring a methodology sends the signal that the organization values this way of working. That creates an incentive to use design thinking. Additionally, when we begin to compare teams' process to the outcomes we can begin to describe the real impact of design thinking in organizations. In essence, we will be able to respond to the question, do teams that have a strong process produce more innovative outcomes?

References

Amabile TM (1996) Creativity in context: update to "The Social Psychology of Creativity". Westview Press, Boulder

Bandura A (1977) Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 84 (2):191

Bandura A (1994) Self-efficacy. Wiley, New York

Barab S, Squire K (2004) Design-based research: putting a stake in the ground. J Learn Sci 13 (1):1–14

Courage C. (2013) Reweaving corporate DNA: building a culture of design thinking at Citrix. Resource document. Management Innovation eXchange. http://www.managementexchange.com/story/reweaving-corporate-dna-building-culture-design-thinking-citrix. Accessed 12 Dec 2013

Cross N (1990) The nature and nurture of design ability. Des Stud 11(3):127–140

Dow SP, Glassco A, Kass J, Schwarz M, Schwartz DL, Klemmer SR (2010) Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. ACM Trans Comput Hum Interact 17(4):18

Dow SP, Klemmer SR (2011) The efficacy of prototyping under time constraints. In: Design thinking. Springer, Heidelberg, pp 111–128

Dweck CS (2000) Self-theories: their role in motivation, personality, and development. Psychology Press, Philadelphia

Dweck C (2006) Mindset: the new psychology of success. Random House, New York

Glaser BG, Strauss AL (2009) The discovery of grounded theory: strategies for qualitative research. Transaction Publishers, New Jersey

Guilford JP (1968) Intelligence, creativity, and their educational implications. RR Knapp, San Diego

Hawthorne G, Quintin EM, Saggar M, Bott N, Keinitz E, Liu N et al (2014) Impact and sustainability of creative capacity building: the cognitive, behavioral, and neural correlates of increasing creative capacity. In: Design thinking research. Springer, Heidelberg, pp 65–77

Kelley T, Kelley D (2013) Creative confidence: unleashing the creative potential within us all. Random House, New York

Marelaro 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. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer, Heidelberg, pp 41–57

Nussbaum B (2011) Design thinking is a failed experiment. So what's next? Fast Company, April. Retrieved from http://www.fastcodesign.com/1663558/design-thinking-is-a-failed-experiment-so-whats-next

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. Springer, Heidelberg, pp 95–105

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

Royalty A, Ladenheim K, Roth B (2015) Assessing the development of design thinking: from training to organizational application. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Springer, Heidelberg, pp 73–86
Runco MA (1999) Divergent thinking. In: Runco MA, Pritzker SR (eds) Encyclopedia of creativity, vol 1. Academic, San Diego, pp 577–582
Schunk DH (1991) Self-efficacy and academic motivation. Educ Psychol 26(3–4):207–231

Part IV
Documentation and Information Transfer in Design Thinking Processes

Experience and Knowledge Transfer Through Special Topic Coaching Sessions

Franziska Häger, Thomas Kowark, and Matthias Uflacker

Abstract Design teams have different problems and needs during their projects. For the teams' coaches this can amount to a high workload due to acquainting themselves with new topics and preparing different coaching sessions for the teams. On the other hand, former teams might have experienced similar problems, and the experiences and solutions of members from such a team could be valuable to the current teams. In this chapter, we present the concept of special topic coaches that lets former members of design teams provide additional coaching sessions to current teams. This concept allows experienced members of design teams to share their knowledge and experiences with a current team by preparing and running a coaching session. Thus it relieves the dedicated team coaches while providing the team with valuable input and establishing an exchange between different design teams. The teams can discuss their problems with peers and receive additional input on topics important to them when needed. To ensure that our goals are met, we implemented the concept with the help of a coaching seminar parallel to one of our design engineering courses. The seminar was evaluated through observations and questionnaires. We will present the results as well as derived changes to the next iteration of the seminar and our ideas of how to adopt this concept in a company context.

1 Introduction

Former members of design teams are often used as coaches for new teams to pass on their knowledge and previous experiences. This can be seen in design courses as well as in companies implementing design and innovation teams. However not all members of a design team will become design thinking coaches later on, which often means their expertise and experiences are unreachable for current and upcoming design teams.

In this chapter, we present a concept that allows former members of global design teams to prepare and execute special topic coaching sessions for current

F. Häger (✉) • T. Kowark • M. Uflacker
Hasso Plattner Institute for Software Systems Engineering, Campus Griebnitzsee, 14482 Potsdam, Germany
e-mail: franziska.haeger@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_13

teams. These special topic coaches also have to prepare lean documentation about the respective topics in order to allow for future reuse. The coaches offer a variety of topics (e.g. decision making, interview acquisition, or presentation preparation), which are assigned to the teams in accordance with the project schedule and their current demand.

With this concept we aim to offer teams additional helpful coaching sessions, while taking some of the workload from the teams' main coaches. Additionally we hope to improve the knowledge exchange between current and former teams and thus establish a stronger community of design team members in general.

We implemented our concept in the form of a seminar in which alumni of our global design and engineering course (Carleton and Leifer 2009) take on the role of the special topic coaches. To evaluate the quality and impact of their coaching sessions, we qualitatively analyzed each coaching input by observing the dynamics of the session and identifying positive and negative aspects of coaching work. Additionally, we performed a quantitative analysis by using questionnaires that assess how well the input was perceived by the team, how helpful it was at the respective point in the project, or how the ratio between practice and theory was perceived by the team. The coaching sessions are observed with minimal interference (i.e., observers only interfere in case of wrong inputs) in order to allow for an undistorted assessment of their quality.

The rest of this paper is structured as follows: Sect. 2 reflects on related work in coaching and knowledge management, Sect. 3 introduces the general concept of special topic coaches and implementation requirements, Sect. 4 describes our implementation of the concept in form of a parallel coaching seminar to our global design course. The evaluation of the seminar and the coaching sessions is discussed in Sect. 5 while Sect. 6 gives an outlook on the next iteration of the course and concludes the paper with a summary.

2 Related Work

Coaching is a common concept in innovation, design, and engineering teams. This holds true for educational as well as workplace settings, as shown by the courses described in Carleton and Leifer (2009), Gerber et al. (2012), and Ledsome and Dowlen (2007), or as shown in case studies and experience reports like Meyer and Harling (2012) and Hiremath and Sathiyam (2013). Additionally, a related study on future trends in workplace learning settings by Kim et al., found that coaching and mentoring are important strategies in blended learning approaches (Kim et al. 2006).

Reich et al. have created a conceptual framework identifying five different coaching styles known as: the consultant, the supervisor, the instructor, the facilitator, and the mentor. The styles differ in terms of the relation between coach and team, the goals of the coach and the way the coach intervenes (Reich et al. 2009). Building on this work, Reich et al. evaluate the perception of these coaching styles with students, coaches and teaching staff (Reich et al. 2007). They found that coaches rated roles with a higher workload for themselves as worse than students did, e.g. the instructor, facilitator, and supervisor roles. Furthermore they found that high-performing teams and coaches saw their main roles as being a mentor, facilitator and supervisor while coaches of low-performing teams saw themselves mainly as mentor. These findings indicate that teams have different coaching needs and require coaches to follow more than one coaching style as needed. If the team-dedicated coach has to fulfill all the coaching needs of his team he will have a higher workload. We address this issue with our concept of special topic coaches by introducing additional coaches, thus lowering the workload of the dedicated team coach and satisfying the coaching needs of the team.

Hackman and Wageman found that effective coaching requires four conditions to be met, two of which have to do with the coaches' actions (Hackman and Wageman 2005):

- "Coaching behaviors focus on salient task performance processes rather than on members' interpersonal relationships or on processes that are not under the team's control", and
- "Coaching interventions are made at times when the team is ready for them and able to deal with them"

We disagree with the first condition, as our teams specifically asked for help with interpersonal issues, especially with cultural difficulties. A topic that has also been addressed by other researchers, e.g., Grimheden et al. introduced special culture coaches who served as ambassadors for the "other" culture (Grimheden et al. 2006) in a global virtual team. In our seminar one of the alumni coaches gave a coaching session addressing the cultural difficulties while the global team was visiting. We agree with the second condition and want to address it by giving additional coaching sessions when the teams request them.

Inter-project knowledge management is a complex topic that faces specific challenges (Pretorius and Steyn 2005; Shinoda et al. 2013) due to the unique and temporal nature of projects. Shinoda et al. list the following challenges for knowledge management in project oriented environments (Shinoda et al. 2013):

1. Retention difficulty—knowledge is fragmented due to the routines and structure in projects, and thus projects struggle to retain ideas
2. Inefficiency—if previously generated knowledge is unavailable solutions have to be created again
3. Low replication—projects have higher learning curves due to their uniqueness and low repetition
4. New relationships—people work as a team for a time and then separate, which makes continuous learning difficult

Pretorius and Steyn have identified two strategies for knowledge management in such environments (Pretorius and Steyn 2005):

1. Codification—storing explicit knowledge in a database
2. Personalization—sharing (tacit) knowledge person-to-person

They also state that it is important to share tacit knowledge through human interaction.

In our annual course we have seen some of the identified challenges and we believe that the concept of special topic coaches can help overcome these by providing a personal exchange of such knowledge as proposed by Pretorius and Steyn.

3 The Concept of Special Topic Coaches

As previously described, different design teams have different needs with regard to coaching. While dedicated team coaches are well prepared to give general input on design methodologies the teams often require additional input on special topics that arise during the project and are often project specific. Expecting the dedicated team coaches to prepare coaching sessions on all these topics would require them to get acquainted with completely new topics or problems in which they do not yet have expertise. Additionally, preparing such sessions results in a much higher workload. On the other hand, members of other design teams can be experts on the topic or have encountered similar issues in their own projects.

These facts have led us to create the concept of special topic coaches. The concept introduces current design teams to a pool of experienced design team members who are willing to provide inputs and coaching sessions on very specific topic they are experienced in.

We believe that by introducing members of other or former design teams as coaches for a specific topic we can foster exchange between projects, relieve the dedicated team coaches from some of their workload and provide design teams with valuable input for their current needs. To achieve such a setup it is necessary to have an overview of available coaches and their special topics and to make the design teams aware of the possibility of getting coaching for special topics/issues.

The first goal can be achieved, e.g. by creating profiles of people who are willing to give such coaching sessions including their areas of expertise as well as special issues and characteristics of the projects they have worked on. Such profiles can for example be collected by the dedicated team coaches at the end of a project and then be kept in an intranet or a social platform, or they could be managed by a special coaches coordinator.

To make the teams aware of the special coaching topics, the dedicated team coach should use reflective sessions. This could mean, for example, finding out where additional input is needed and which open issues exist at the end of a day or a working phase. The team coach or a coordinator for special coaches could then make a connection between the team and an appropriate coach, to organize a special coaching session.

Another important aspect to consider is how to introduce these special coaches to general coaching practices. This can for example be achieved through workshops or trainings with experienced coaches.

To create a valuable coaching session, the coach should take some time with the team he will coach to learn about their project and their current problems and needs. He can then prepare and implement one or more coaching sessions for and with the team. In order to document the coaching for reuse by himself or other coaches the coach should also create a lean documentation of his work.

4 Coaching Seminar for ME310

In order to evaluate our idea of special topic coaches, we implemented a coaching seminar for alumni of a graduate-level course This course is part of a global design education effort that originated from Stanford's ME310 course (Carleton and Leifer 2009).

It is a 9 month, global, project-based, design course in which a design team consists of students from two or three universities. Each team works on a challenge given by an industry partner. The teams are supported by a corporate liaison from the industry partner and a team coach, usually recruited from course alumni. In addition, teachers and coaches give inputs and advice to all teams. Figure 1 illustrates the timeline of the course.

The 9 months are structured into different assignments focusing on the exploration of user needs, prototyping and testing. The first prototypes focus on exploring different solutions. Starting with the Functional System Prototype the assignments focus on implementing the final idea.

Similar to other universities in the network that have formed around the ME310 course, we found that employing alumni as permanent team coaches is a valuable approach. In this way former students can share their experiences with the new team

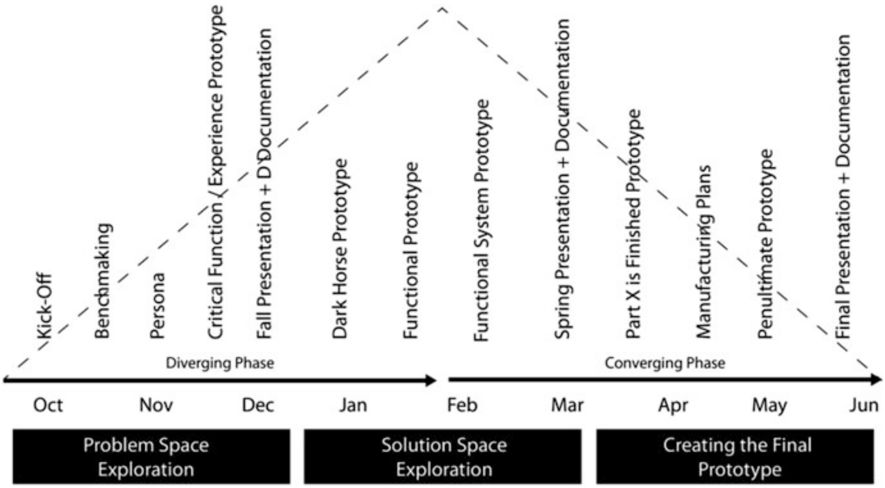


Fig. 1 Timeline of the global design course

and have the possibility to gain insight into teaching and coaching. Therefore each of our design teams in ME310 has a dedicated alumni team coach.

The parallel coaching seminar which implements our concept of special topic coaches runs during the first two thirds of the global design course. It allows additional alumni of the course to gain experiences in coaching and share their knowledge on different topics. It also provides the current project teams with a possibility to receive coaching on specific topics as needed and takes some of the workload of the dedicated coaches. The seminar thus addresses the issues we hope to solve with the concept.

Figure 2 depicts the sequence of events of the seminar. In a first meeting, all alumni coaches were given an introduction to general coaching guidelines and the advisors discussed possible topics with each alumnus based on their expertise and experiences in their own projects. In addition, the advisors spoke with the three current design teams and discussed the issues in which they wanted to receive special coaching sessions. Based on these possible topic collections the advisors assigned each alumnus to one of the three teams. The coaches met with the team they were to coach and discussed the final topic and the team's needs with the students. Afterwards they prepared a coaching session and a lean documentation on the topic chosen together with the team. This lean documentation served two purposes. First, it was a handout for the coached team to remember the most important aspects of the session. Second, it allowed the dedicated team coaches of the other teams to repeat the coaching session with their team if the topic should be interesting to them. The concept and the documentation were then reviewed with the seminar advisors, before actually coaching the team. During the actual coaching session seminar advisors took part as silent observers.

To evaluate the concept in general, the coaching efforts and the impact on the teams we devised a concept of evaluating each coaching session. This concept

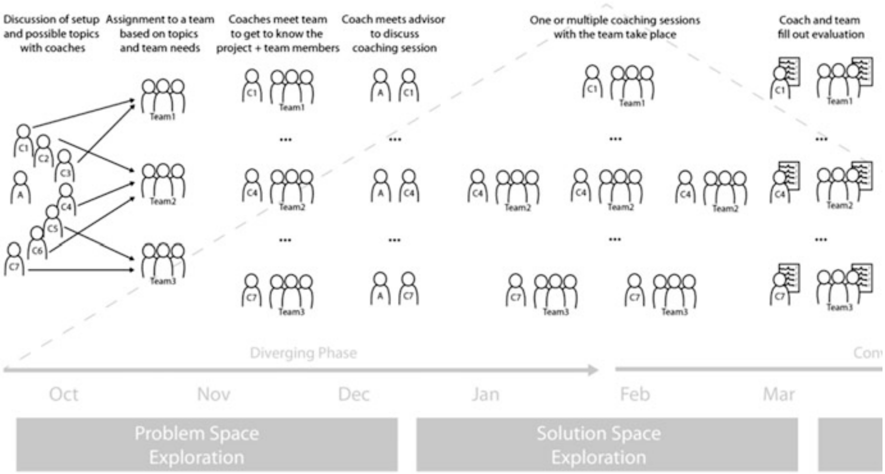


Fig. 2 Timeline of the coaching seminar

allows us to gain additional insights into the dynamics, and positive and negative aspects of such coaching sessions. The evaluation concept consists of two questionnaires, one for coaches and one for the coached students and an evaluation sheet for the teaching staff. Alumni coaches as well as the coached teams were asked to fill out the questionnaires after their coaching session, rating the session in general, the use of tools (e.g. handouts, whiteboards, and exercises), the ratio of theory vs. practice, and pointing out positive and negative aspects of the coaching sessions. The advisors filled out the evaluation sheet during the session, rating preparation, structure, theoretical and practical components, empathy of the coach for the team and their general impressions.

5 Evaluation

In this year's iteration of the seminar, seven course alumni gave coaching sessions. All the alumni were first time coaches and alumni of last year's course. Each coach gave a session to one of the three four-student teams. For one of the coaching sessions the partner team was available so the session included seven students. The topics coached included decision-making as a team process, interview acquisition, global documentation, presenting abstract concept in a tangible manner or how to receive honest user feedback. Depending on the topic, the coaches gave one or multiple input sessions lasting between 30 min and 4 h and provided two to four pages of documentation. The first results show that alumni value the possibility of passing on their gathered knowledge by giving coaching sessions, while teams find alumni input valuable for their ongoing projects.

5.1 Self-Evaluation of the Coaches

Six out of seven coaches answered the questionnaire after giving their coaching session. We asked the coaches to describe positive and negative aspects of their coaching session, the tools they used and whether they found them helpful, and the ratio between theory and practice.

5.1.1 General Impressions

The positive aspects of the sessions, as described by the coaches, included engaged team members, and practical tips and personal stories from the coaches own projects as a good way to explain their recommendations. The main disadvantages mentioned were the limited time of the team members and some issues with executing their coaching ideas, for example: "I wish I had more experience with regards to the balancing act between the actual prototype and the requirement of the cooperate partner." or "Maybe we could have done a warm-up exercise or anything similar."

5.1.2 Tools

When asked about tools that were used during the coaching session most coaches mentioned the whiteboard, the handout and/or slides, and the timer. Three of the coaches found that the whiteboard was useful to the collaborative work of the team as the following statement describes: "The whiteboard was really useful in discussions and ideation." Two of the coaches found the documentation in the form of handouts or slides useful for themselves as a guide but not so much for the team. One coach described this as follows: "[The] handout was only helpful for myself, not really for the team at the time of the coaching session." In contrast, other coaches found the documentation very helpful for the team as a summary and a reminder for the future, for example: "I think the most helpful artifact, that resulted from the coaching is the handout I prepared for the team members. By having all important information as a rule set at hand, it will be easier to remember and act on the advice given in the next user testing." It can be argued that the difference here depends on the amount of theoretical information needed for some topics as opposed to other rather practical topics where team and coaches mostly worked together and a summarizing document is less needed. The timer was only mentioned by two coaches and appeared only important for sessions with a schedule of inputs and exercises and not so much for sessions mostly based on discussions.

Interestingly, most coaches described tools they used with the team during the coaching session. However one of the coaches described a tool he used by himself alone: "I essentially prepared a full discussion agenda with all points that I would want to bring across plus possible responses on the team's side that I could use to follow, skip parts that were not applicable or deep dive into others they were interested in. That was very helpful to keep track of the content, to not diverge too much and to always know what to say next." This idea seems very helpful for topics that involve longer discussions.

5.1.3 Theory vs. Practice

Naturally, different coaching topics require a different level of theory and practice. When asked to rate the ratio between theory and practice in their coaching session most coaches described the level as corresponding to the topic, as the following statements reflect: "The topic being very theoretical in nature, there was only limited practical coaching to do. I saw no need for practical coaching for this topic." and "The team was already aware of the goal of the theory, so I needed less time to introduce it to them and we were able to focus more on the prototype itself."

5.1.4 Summary

Overall the alumni valued the experience of coaching. One coach stated “I really like the opportunity to actual[ly] coach, which is still rare in university in general.” Another coach wrote “I like the idea of a coaching seminar. I also like, that the ME310 alumni are still part of the ME310 family, with the coaching seminar more than ever.” This highlights, that the seminar provides an opportunity for the alumni to stay connected with the course creating more of a community between alumni and current students.

5.2 Evaluation of Coaches by Student Teams

During the global design course each of the three teams was coached by two or three alumni coaches. One coaching session even involved students from the global university who were visiting during the time of coaching. After each session, all participating students were asked to fill out the questionnaire. We received 21 of 31 possible responses. For each coaching session at least two students answered the questionnaire. Similar to the coaches, the students were asked to describe positive and negative aspects of the coaching session, tools used and whether the tools were useful and the ratio of theory and practice during the coaching session.

5.2.1 General Impressions

The major positive aspect the students saw in the coaching sessions included obtaining input on a topic to help them with a current problem or need. For example, students said: “(the session) structured our process of finding interview partners,” “useful coaching which helped us to explain our project topic better” or “we gained really good insights into testing especially with elderly people.” Negative aspects were found mostly in the timing of the coaching sessions as most students wished to have had the sessions earlier in the project. This can be seen in the following reflections “I would have liked the coaching a little earlier, as some points are really valuable … especially pointing out some points again, that we forgot over the time.” or “the whole coaching session was maybe a little too late in our process and we couldn’t use as much as we would have liked to” This lateness is due to the fact that the global design course starts about a month earlier than other courses at our university, while coaching commenced at the beginning of the semester.

5.2.2 Tools

The student's evaluation of tools used during the coaching session matches the coaches' evaluation. The whiteboard and the documentation in the form of handouts or slides were most mentioned. The students said: "structuring [the] concept on whiteboard was quite helpful" and "visualizing the different priorities for our project [on the whiteboard] was really nice for understanding the team." Students therefore agree with coaches that whiteboards are a good tool for collaborative coaching sessions. The handout was described as "pretty helpful" or "very helpful for future reference". Unlike the coaches, no student experienced the handout as not being helpful. While the timer was mentioned as a tool in use, no comments were made on its usefulness. This is probably because students don't realize the importance of timing in cases where the coach has a good time management.

5.2.3 Theory vs. Practice

Similar to the coaches, the students found the ratio between theory and practice well-suited for the respective topic as reflected by statements like "[It was] a good mixture between exercises and theory." or "[The coaching was] very practical, which is good for this topic because I think there is not too much theory."

5.2.4 Summary

Overall the students valued coaching from course alumni highly as it gave them insight into former projects and provided them with additional input when they needed it. For example, "[the coaching] was in the perfect project phase" or "the general theme was awesome as we [as a team] have problems with it".

5.3 Coach and Student Ratings

Complementary to the text based questions, students and coaches were asked to rate the coaching session in general and the ratio between theory and practice on a Likert Scale ranging from 1—"Very good" to 5—"Not good at all". Additionally, the students were asked to rate how well the coaching session fit into their current process.

5.3.1 General Ratings

Figure 3 depicts the general student (S) and coach (C) ratings for the seven coaching session in a box-and-whisker plot. As can be seen, all coaches rated their session inside the range of student ratings. The rating of the coaches from coaching session

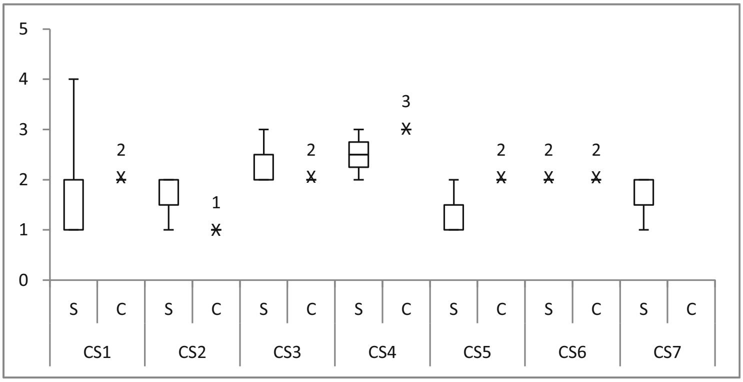


Fig. 3 General ratings of coaching sessions by coaches and students

CS1, CS3 and CS6 are within the second and third quartile of student ratings. The rating of the coaches from coaching session CS2, CS4 and CS5 correspond with the outliers of student ratings. For these coaches it would be recommended to reflect on the students' textual answers to see why most of them rated the session better or worse than the coaches themselves.

5.3.2 Rating of Theory and Practice

Figure 4 shows the ratings of the theory-to-practice ratio of students and coaches for the seven coaching sessions. As can be seen, coach and student ratings differ more on this topic as only the coach rating from session CS4 is between the second and third quartile. The coaches from sessions CS2, and CS5 correspond with the worse outliers of student ratings. And the coaches from sessions CS1 and CS6 rated the ratio worse than any of the students. In CS1 the coach would have liked to have found some more time for theoretical inputs, while the coach from CS6 would have liked to have had more time for a practical session with the team. Only the coach from session CS3 rated the ratio better than most of the students. It would be interesting to look further into the different ratings on this topic by interviewing students and coaches again. However, we think the ratings and the associated comments by the coaches and students indicate that the coaches have higher expectations towards their coaching session and can't fit everything they would like into the time slots for coaching sessions. Conversely, the students find the provided input sufficient.

5.3.3 Rating of Suitability to Process

Figure 5 depicts the rating by the students of how well the coaching session fit into their current process. As can be seen, the student ratings are the worst here. As the comments in Sect. 5.2.1 describe, this is due to the fact that most students would

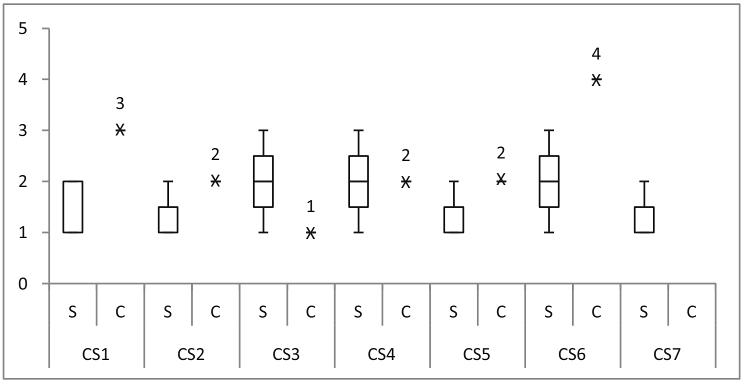


Fig. 4 Theory vs. practice rating of coaching sessions by coaches and students

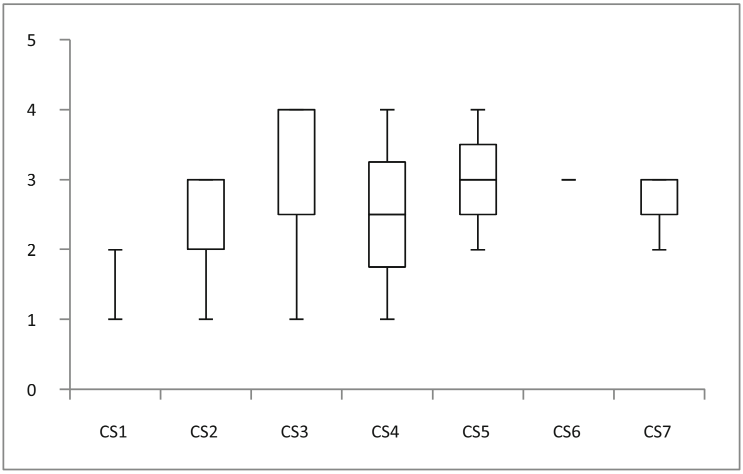


Fig. 5 Ratings for project suitability of the coaching sessions by students

have liked the coaching sessions to take place earlier in the project in order to make them even more useful.

5.4 Evaluation of Coaches by the Advisors

As described in Sect. 4 each coaching session was observed by one of the course advisors. This was done for three reasons:

- To be able to grade the coaching session,
- To be able to intervene in case the coach is giving wrong input, and
- To evaluate the quality of the coaching sessions.

Overall we found that all coaches were able to gain empathy for the team from their preparation session and managed to provide a coaching session with a useful topic for the team. This is also reflected by the coaches and students evaluation. Similar to the coaches themselves and the students we found the ratio of theory and practice fitting for the topics. However, for some of the more theoretical topics, a few exercises could have helped to make the information more tangible, for example, writing a mail to request interviews. Additionally, we noticed some issues in guiding discussions where it seemed the alumni coaches were missing experience in such situations. For example, in one coaching session the students were discussing the same arguments over and over while the coach was just standing there listening and the advisors had to step in to end the pointless discussion. In a different session the students seemed very tired and the coach had a hard time getting them to participate instead of just listening to him. In this session the advisor did not intervene but noted that providing a set of warm-ups and similar tools would be helpful for the alumni coaches.

6 Outlook and Summary

Overall the seminar was perceived positively by teams and coaches. The inputs provided by the alumni gave students the opportunity to discuss their project and problems with peers. Additionally, the inputs were perceived as valuable and relevant as they were mostly aligned with the team's process and gave firsthand insights into former projects. The seminar provided the coaches an opportunity to reflect on their own projects and pass on the knowledge gained therein. Furthermore the seminar made the coaches feel they are still part of the community even though they are not part of the design course as such. Thus the concept of the special topic coaches achieved what we had hoped for in our first implementation.

The positive feedback encourages us to continue this seminar in the future. However, the evaluation uncovered some flaws in our execution, so we plan on some adaptations to the course for the next iteration.

We noticed a couple of actions we would like to collect as "best practices" for a rerun of the seminar. These include:

- Taking some time to get to know the team's current status and problems upfront, e.g. through a personal meeting with the team or reading the latest documentation.
- Starting with a warm-up if the team seems to be sleepy or distracted.
- Discussing the agenda for the coaching session at the beginning with the team.
- Ending a coaching session with action items for the next coaching or team session.
- Incorporate small exercises for theoretical inputs where possible.

We agree with the coaches' assessment, expressing their unpreparedness for some coaching situations, such as long running discussions or dealing with a sleepy and distracted team. Therefore we will extend the introduction of coaching guidelines, for example by creating a kick-off workshop with exercises to help them.

Furthermore students and coaches found that some of the topics were coached too late. In the upcoming project year we will start the coaching seminar in parallel with the global design course allowing the coaching session to take place earlier when the teams most need them.

As discussed in Sect. 5.3.2 the coaches' rating of the theory to practice ratio indicates that they had higher expectations towards their coaching sessions, namely giving more theoretical or practical input than could be achieved in the timeslot. We plan to address this issue by allowing the coaches to give a more detailed theoretical input to all teams and following up with a practical coaching session.

One of the coaches would like to offer a similar seminar for other design courses at our chair. Our approach has proven valuable to students and coaches and helped to enhance the global project course with additional coaching sessions. The teaching staff could not have managed so many individual coaching sessions on their own. Therefore, we firmly believe such a seminar can be valuable for further design and engineering courses at our or other universities.

We also believe that the concept of special topic coaches can be applied to project-based design teams in companies. There it can help to support knowledge sharing between teams and projects and create a stronger sense of community between former and current team members.

References

Carleton T, Leifer L (2009) Stanford's ME310 course as an evolution of engineering design. In: Proceedings of the 19th CIRP design conference—competitive design, Cranfield University, 30–31 Mar 2009

Gerber E, Olson J, Komarek R (2012) Extracurricular design-based learning: preparing students for careers in innovation. Int J Eng Educ 28(2):317

Grimheden M, Van der Loos M, Chen H, Cannon D, Leifer L (2006) Culture coaching: a model for facilitating globally distributed collaborative work. In: Frontiers in education conference, 36th annual, pp 15–20

Hackman J, Wageman R (2005) A theory of team coaching. Acad Manag Rev 30(2):269–287

Hiremath M, Sathiyam V (2013) Fast train to DT: a practical guide to coach design thinking in software industry. In: INTERACT (3), pp 780–787

Kim K-J, Bonk C, Teng Y-T, Zeng T, Oh E (2006) Future trends of blended learning in workplace learning settings across different cultures. In: 29th annual proceedings, Dallas, TX, vol 1, pp 176–183

Ledsome C, Dowlen C (2007) Design coaching. Presented at the international conference on engineering and product design education, Northumbria University, Newcastle Upon Tyne, UK

Meyer J, Harling F (2012) Die harten Nüsse anders knacken: Design Thinking als Innovationsmethode für Wissensmanagement. In: KnowTech—Neue Horizonte für das Unternehmenswissen—Social Media, Collaboration, Mobility, Germany, pp 71–79

Pretorius CJ, Steyn H (2005) Knowledge management in project environments. S Afr J Bus Manag 36(3):41–50

Reich Y, Ullman G, Van der Loos M, Leifer L (2007) Perceptions of coaching in product development teams. In: CDROM proceedings of the 16th international conference on engineering design (ICED), The Design Society

Reich Y, Ullman G, Van der Loos M, Leifer L (2009) Coaching product development teams: a conceptual foundation for empirical studies. Res Eng Des 19(4):205–222

Shinoda AM, Maximiano AA, Sbragia R (2013) Knowledge management in project-oriented organizations. In: BALAS annual conference, Universidad ESAN, Lima, Peru

Smart Documentation with Tele-Board MED

Julia P.A. von Thienen, Anja Perlich, Johannes Eschrig, and Christoph Meinel

Abstract Documentation is a field of active research both for the community of design thinkers and medical practitioners. One major challenge is to combine the advantages of analogue and digital documentation. Since documentation needs are particularly similar in the fields of design thinking and behaviour psychotherapy, an intense collaboration has emerged among these disciplines. The design thinking tool Tele-Board has been adapted for documentation purposes in behaviour psychotherapy, yielding a first medical application of the new tool Tele-Board MED. In the course of tool adaptation, additional features have been developed such as an automatic protocol function. These new features are not only useful for therapists but beneficial for design thinkers too. This chapter explains why collaborative work on documentation tools is particularly promising at the intersection of design thinking and behaviour psychotherapy. We outline the development and empirical evaluation of the new protocol function. Furthermore, this chapter discusses how Tele-Board MED supports documentation in behaviour psychotherapy on the one hand and documentation of design thinking projects on the other hand.

1 Design Thinkers and Behaviour Psychotherapists: A Promising Collaboration

For many communities, work documentation is an important but challenging objective. The communities of design thinkers and medical practitioners are both familiar with a clash of preferences. On the one hand, analogue documentation strategies, such as handwritten notes, are quick and easy at first, when information is being captured. On the other hand, computerized digital documentation has many advantages as well such as automated search functions when information is processed (Leiner et al. 2009; Miller and Sim 2004).

J.P.A. von Thienen (✉) • A. Perlich • J. Eschrig • C. Meinel
Hasso-Plattner-Institute, University of Potsdam, Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, Germany
e-mail: Julia.vonThienen@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_14

The design thinking community has developed a documentation tool to combine the advantages of analogue and digital documentation: the digital whiteboard system Tele-Board (Gumienny et al. 2011, 2012; Gericke et al. 2012). Capturing information can be as easy and quick as handwriting with Tele-Board, e.g., by using a digital pen. At the same time, processing information is convenient due to digital functionalities like automated search or easy copying.

The project Tele-Board MED was launched in 2012 to put the documentation solution Tele-Board and several other benefits of the design thinking approach at the disposal of medical practitioners (von Thienen and Meinel 2012). Due to strongly overlapping documentation challenges, behaviour psychotherapy was chosen as the first domain of intense collaboration. Tele-Board has been adapted to suit the special needs of behaviour psychotherapies, resulting in a first version of Tele-Board MED (von Thienen et al. 2014a; Perlich et al. 2014).

This section will start with a review of major documentation needs and standard documentation solutions in design thinking (Sect. 1.1) versus behaviour psychotherapy (Sect. 1.2). Such a comparison shows that both communities share many needs and already use many similar documentation solutions. Thus, it can be expected that a transfer of documentation tools can take place smoothly and with immediate benefit. Against this background, Sect. 1.3 introduces *Tele-Board MED for Behaviour Psychotherapy* and explains how Tele-Board has been adapted to suit the specific needs of behaviour therapists and their patients.

Section 2 introduces major advances of the Tele-Board MED documenting abilities since an automatic protocol function has been newly developed. Section 3 discusses standard documentation scenarios in behaviour psychotherapy where Tele-Board MED can be of considerable assistance. Section 4 surveys several other design thinking documentation tools and discusses how design thinkers can profit from the new documentation features in Tele-Board (MED).

1.1 Documentation Needs and Solutions in Design Thinking

Design thinkers create human-centred product or service innovations by concentrating on human needs, seeking face to face exchange and conducting playful experiments. Design thinkers are anything but bureaucratic documenters. However, they do need to document their project work somehow. Here are some typical situations:

Touching upon private issues when talking to "strangers" In order to understand potential users, design thinkers often talk to people who belong to the intended user group. While these people are often complete strangers at first, design thinkers commonly want to understand their world view, their needs and their emotions—thus very personal issues (d.school 2011, p. 10). While trying to build up and maintain an empathic relationship with the interviewees, design thinkers also need to document what they find most memorable. Since a bureaucratic documentation scenario (e.g., typing on a computer) might destroy the empathic rapport, design thinkers typically write down just short bits of crucial information, such as striking user statements or observations.

Cooperative information sharing using whiteboards Design thinkers typically work in teams. When they split up—to interview users, for instance—they soon come together again to exchange what they have learned. Typically, these are the circumstances of information exchange: (a) Time is short. (b) There will be a lot of information. It should be easily accessible later on. (c) The team does not know in advance what information they will end up with and what they will do with it subsequently. The typical documentation solution is a whiteboard with sticky notes and sketches. This solution has proven very useful given the circumstances just mentioned. (a) Providing just the most relevant bits of information on sticky notes is a very fast way of information exchange. (b) As the whiteboard is large enough for many sticky notes, it can display a lot of information at one time. Clustering sticky notes and adding sketches eases orientation. (c) When the team thinks about patterns in the information, sticky notes can be arranged accordingly on the whiteboard. By way of contrast, the notes could not be handled as flexibly if the team had written down everything on a single sheet of paper. In that case, they would have to cut out the notes first to move them around and build new clusters. Or they would have to write down everything anew in the preferred arrangement.

Justifying work to externals by demonstrating exciting work results Design thinkers do not only report to team-mates. They also report to externals every once in a while, in particular to the initiators of design thinking challenges such as business companies who fund the projects. When communicating to externals, teams typically demonstrate their work progress by illustrating basic user needs they identified, by providing a problem analysis (why or how users currently fail to meet their needs) and laying out the envisioned solution. In contrast, design thinkers usually do not hand over bureaucratic protocols of actions they took to justify the money they received from funding partners.

Seeking experiences away from the whiteboard Throughout their project work, design thinkers often leave their "base station" of documentation—the team space with whiteboards—to meet people or try things out. However, project work away from the whiteboard needs to be documented too, at least in a rudimentary fashion. Often-times, pictures are taken that can be placed on the whiteboard later on. Sometimes, videos are recorded. Such visuals also help to capture what words can barely convey: atmosphere, mood, the background setting etc.

Analysing by promoting common sense Design thinkers dig deep into the information they gathered. They want to understand the core of a problem and discern basic unsatisfied needs. To arrive at sharp-witted insights, design thinkers use open, human-centred analysis schemes that people fill in with a good deal of common sense and even creativity. There are *point-of-view* madlibs to pinpoint the core of a problem (d.school 2011, p. 21). There are *I-wish-I-like* templates to capture the pros and cons of experiences (d.school 2011, p. 44). There are *journey* *maps* to grasp the central theme in historical developments (d.school 2011, p. 16). However, there are no mathematical deductions, no formal theories, no category systems—all in line with the mottos “structure without formalizing” and “focus without formalizing” (von Thienen et al. 2014b, p. 102). Human stories should always be tangible, even in the most sharp-witted analyses.

Traditionally, design thinkers used analogue whiteboards and sticky notes made out of paper as central documentation tools. However, a digital whiteboard software application called *Tele-Board* has been newly created for design thinking project work (Gumienny et al. 2011, 2012; Gericke et al. 2012). It brings along many advantages. In the past, teams had to clean their whiteboards regularly to make room for new pieces of information. Tele-Board, however, saves all stages of progress on a whiteboard. This way, teams don’t lose their information overview from an initial exploration phase when they move on to ideation (thinking about potential solutions), prototyping or testing. Furthermore, when teams work with analogue whiteboards and sticky notes, most information is thrown away at the end of projects. Tele-Board, however, saves project documents without consuming extra space physically—both while projects are running as well as when they are finished. Finally, analogue sticky notes defy automated search, copying or electronic posting. Because content is stored digitally in Tele-Board, teams can easily search for pieces of information automatically, export information to text or picture files, copy information and send it anywhere in the world they want. Thus, Tele-Board helps design thinkers enjoy all the advantages of digital work. At the same time, the system is tailored to create a work feeling that resembles analogue design thinking work as much as possible, for example a digital pen can be used instead of having to type. The work flow, hand-movements or eye-contact—everything which feels good in the process of documenting with analogue tools should feel just as good when documenting digitally with Tele-Board.

1.2 Documentation Needs and Solutions in Behaviour Psychotherapy

Behaviour psychotherapists share most of the needs that design thinkers have and they already apply many similar solutions too. However, there are a few differences.

Touching upon private issues when talking to “strangers” Just like design thinkers who try to establish and maintain an empathic rapport with interviewees, psychotherapists seek an empathic rapport with patients. In both cases, the flow of private information is rather one-directional, i.e. from patient to therapist in the medical context and from interviewee to interviewer in the context of design thinking. Furthermore, in both cases the person who is being interviewed typically has no personal relationship with the interviewer. In this sense it is an information

Fig. 1 It is not uncommon for behaviour psychotherapeutic practices to provide analogue whiteboards so that therapist and patient can write something down cooperatively. Thus, switching to digital whiteboards with Tele-Board MED does not call for a complete change of therapeutic habits

exchange between "strangers". Thus, the background scenario of psychotherapists and design thinkers is very similar when it comes to the aspect of documentation. While some key observations need to be captured in some way, the documentation procedure must not get in the way of empathic rapport. Thus, it is not surprising that both communities have come up with similar documentation solutions. In both cases only the most essential information is written down, usually by hand.

Cooperative information sharing using whiteboards While there is usually a one-directional flow of private information (from patient to therapist), there are many other therapeutic settings in which a bi-directional flow of information is important. For instance, the therapist might engage in psycho-education telling the patient about a typical vicious cycle of anxiety and avoidance behaviour. Then, the patient adds to the information from the therapist information about his own case. Together, patient and therapist construct a vicious cycle for the patient that he ought to counter in the future.

To allow for such in-therapy collaboration, many therapeutic practices provide analogue whiteboards quite similar to the typical design thinking setup for team-work (Fig. 1).

Thus, once again the documentation needs and solutions resemble each other in behaviour psychotherapy and design thinking.

Justifying work to externals by sending bureaucratic case reports One key difference in documentation needs and solutions concerns the communication with externals. While design thinkers typically meet the business partners who pay for their projects in person, so that information can be exchanged personally, standardisation and anonymity are crucial in the psychotherapeutic context. Therapists send official "bureaucratic" reports to insurance companies in order to obtain payment for treatments. In these reports, therapists describe the cases of patients, treatments planned and perhaps treatments carried out in the past followed by an evaluation of effects.

Seeking experiences away from the whiteboard Just like design thinkers, therapists and patients often spend time away from the whiteboard—in case there is one in a therapeutic practice. Maybe the patient tries autogenic training. In another session, the patient might put down stones and flowers across the room to reconstruct his "life line" including major burdens (stones) or uplifting experiences (flowers). Or they could work with angst-inducing stimuli to reduce avoidance behaviour. In every such case the focus is typically on experiences. The "workflow" has priority over documentation. Detailed notes regarding such sessions are unlikely to emerge. Rather, photos or videos can help capture important content without disturbing the flow.

Analysing by building on psychotherapeutic theories Like design thinkers, therapists invoke schemes to analyse information. However, the schemes used in psychotherapy tend to be less playful. They built on psychotherapeutic theories and are relatively formal.

Ensuring confidentiality of data While design thinkers typically share their findings with broad audiences, therapists barely communicate with anyone regarding the information they obtain. In behaviour psychotherapy, the protection of patient data is at least as important as in all other healthcare domains. The professional code of secrecy states that everybody who works in health services has to keep strictly confident all information obtained concerning the patient (Bundesärztekammer 2008a, b; Leiner et al. 2009). This implies that medical records have to be stored in such a way that they cannot be accessed by third parties. In practice, the access to paper-based records is physically restricted as they are stored in a locker. In digital documentation, electronic measures are essential to protect records from unauthorized access, e.g., through data encryption and password requests.

In sum, design thinkers and behaviour psychotherapists share many documentation needs and they already share many solutions too. In particular, documentation strategies must go hand-in-hand with empathic conversations. There have to be documentation strategies when the information flow is almost entirely one-directional (as a patient or user reveals private experiences), when the information flow is back and forth (such as when patient and therapist carry out an analysis together and design thinkers exchange information on a whiteboard) or when the focus of work is not on documentation at all (because a patient practices autogenic training, for instance, or design thinkers seek first-hand experiences regarding their challenge).

However, there are also three major differences regarding documentation needs and solutions. In particular, data safety is generally a lot more important in the psychotherapeutic setting. Different kinds of analysis schemes are used. Finally, psychotherapists need to generate bureaucratic case reports while design thinkers do not.

1.3 Suggesting a Design Thinking Based Documentation Solution for Behaviour Psychotherapy

Behaviour psychotherapists currently face the challenge of rethinking traditional documentation formats. Therapists need to adhere to new laws that grant patients the right to access all documentation regarding their treatments (e.g., Bundes-Patientenrechtegesetz 2013). Like other medical practitioners, behaviour psychotherapists ask themselves whether they should keep up with general developments in the digital age (von Thienen et al. 2014a). Furthermore, there are strong movements towards patient empowerment. Patients shall obtain better “access to information” (Koch 2012, p. 26), they shall be supported when “building knowledge” (ib.) and “transforming knowledge into action” (ib.). Provider-patient interactions should be rethought and revised. Health care providers ought to share more knowledge with patients and they should grant patients more power to set the course of action regarding their own treatment (Aujoulat et al. 2006).

As behaviour psychotherapists rethink standard documentation formats, it is a likely move to look for other communities that have similar documentation needs and, potentially, some inspiring solutions. This seems to be the case in design thinking. Behaviour psychotherapists and design thinkers share many documentation concerns and both already use many similar solutions (cf. Sects. 1.1 and 1.2). Furthermore, design thinkers are acknowledged team workers. They are known for flat hierarchies and collaboration across cultures and professions (Plattner et al. 2009). Thus, they obviously possess documentation solutions already which support teamwork at eye-level. Finally, in view of Tele-Board, the design thinking community has already started the demanding shift from analogue to digital documentation. After all, digital documentation is attractive in design thinking projects only when many constraints are met—just as in psychotherapy. Documentation must not disturb empathic rapport. It must leave ample room for spontaneity. It must not disturb the flow of experiences when people try something out.

We have therefore set out to adapt Tele-Board for documentation purposes in behaviour psychotherapy—and maybe other medical domains to follow. A first version of Tele-Board MED has been created (Fig. 2).

To obtain Tele-Board MED, the original Tele-Board had to be expanded, reduced and changed in some respects.

Expanding Tele-Board As the direct comparison of documentation needs and solutions showed (cf. Sects. 1.1 and 1.2), there are three major needs in behaviour psychotherapy not shared by design thinkers. Therefore, in these three domains Tele-Board had to be expanded to suit its new field of application: (1) rigorous data security standards are essential; (2) the system needs to support administrative tasks like the writing of case reports for insurance companies and (3) analysis schemes need to be based on behaviour psychotherapeutic theories.

Reducing Tele-Board Some Tele-Board features, such as a means to support remote collaboration based on video conferences, do not seem immediately useful

Fig. 2 Tele-Board MED is a digital whiteboard system for medical documentation. Patient and therapist can use it to document cooperatively. Information can be entered by typing or by using digital pens

in the new field of application. Features that might more likely cause confusion than be of assistance have been deactivated in Tele-Board MED.

Changing Tele-Board The graphical user interface had to be redesigned. Tele-Board MED has to organize patient files whereas Tele-Board organizes design thinking projects. Thus, the terminology of user menus and help files had to be adapted.

While the implementation of rigorous data security standards has been an important step in the creation of Tele-Board MED (Perlich et al. in press), this chapter focuses on documentation concerns. The following sections will discuss how research informed the development of new documentation tools in Tele-Board MED and how Tele-Board MED supports either behaviour psychotherapists or design thinkers.

2 Designing a New Documentation Tool: The Automatic Tele-Board MED Session Protocol

Design thinkers and behaviour psychotherapists need *comprehensive documentation* as much as they need *condensed documentation*.

In *comprehensive documentation*, every detail of a project is recorded; no information gets lost. Thus, every tiny piece of information regarding a project can be looked up later on. However, there is typically so much information, it is difficult to get an overview. Furthermore, finding pertinent information can be time consuming in such large information pools. Therefore, condensed documentation is a helpful complement.

Condensed documentation should provide only the most important pieces of information. Ideally, all the listed information can be surveyed at once. Relevant information is found immediately, search times are close to zero.

Video recordings of complete therapy sessions are an example of comprehensive documentation in behaviour psychotherapy. By contrast, short protocols which name the most important interventions and outcomes of therapy sessions are examples of condensed documentation. Both forms of documentation are presently used in behaviour psychotherapy; both are essential.

The original Tele-Board system supports the first but not the second documentation objective. All digital actions that teams undertake in the Tele-Board system are saved. Whatever keywords teams enter and delete afterwards, however they arrange and rearrange their information—everything can be reconstructed by means of the Tele-Board history browser (Gericke et al. 2012, 2014). Thus, the history browser is an extremely comprehensive documentation tool: It stores and makes accessible basically every detail.

To support the second documentation objective as well—the creation of condensed work summaries—a new and complementary documentation tool had to be added. It should help to overview just the most important steps and findings in a work process. We envision a protocol that is generated automatically or semi-automatically by Tele-Board MED. In this section we outline how such a new documentation tool has been developed, tested and refined for Tele-Board MED.

2.1 Exploring the Status Quo of Protocols in Behaviour Psychotherapy

To gain an initial understanding of present-day therapy protocols, we analysed protocol templates suggested to therapists in training at the BFA (Berliner-Fortbildungsakademie), a training institution for behaviour psychotherapists where about 200 therapists are presently enrolled. We observed over 100 therapy sessions led by three different experienced therapists (each with more than 20 years of work experience) both at the BFA and at an ambulant psychotherapeutic group practice, gaining insight into the notes that the therapists took themselves. Furthermore, we analysed exemplary session protocols that were given out to behaviour therapists in training at the BFA as part of their training material.

2.2 Devising and Refining a New Documentation Concept: The Short Visual Protocol

Building on the design thinking approach of using visuals, i.e. small graphical elements and short verbal notes, a basic structure was developed for Tele-Board

MED session protocols. In terms of verbal content, protocols were supposed to rely solely on utterances made explicitly by the patient or the therapist in a session. Thus, no delicate thought of the therapist and no abstract reflection on the process (e.g., regarding patient resistance) would enter the protocol. Patients could read such protocols without being surprised by their content; everything should sound familiar to them.

This protocol structure was tried out and refined in more than 30 therapy sessions. A member of the Tele-Board MED team created protocols of therapy sessions manually based on the suggested protocol structure. These protocols were then given to the therapists who had led the corresponding therapy sessions to comment on protocol accuracy and utility. After a couple of iterations the basic protocol structure did not change much anymore.

In a feedback study (von Thienen et al. 2014a) the envisioned Tele-Board MED protocol function including a typical end result, i.e. a sample session protocol, was shown to 34 behaviour psychotherapists. Among several suggested new Tele-Board MED features, the therapists considered the protocol function most useful. On a scale of 1 (marginally useful) to 4 (highly useful) it got an average rating of 3.1.

In terms of qualitative feedback, the protocol function was generally considered helpful in remembering prior sessions quickly. However, study participants also addressed two shortcomings. (1) Session protocols should include some things that are not explicitly verbalized in therapy sessions like psychotherapeutic interventions/methods used or the process phase. (2) There should be a way to document things which the patient should not be able to read. In particular, there should be protocols which document problems of the therapist. Such protocols could help to prepare for counselling sessions with experienced therapy supervisors.

Based on this feedback, the general protocol structure was revised once more. It should be easy to include therapeutic methods and the process phase of therapies in Tele-Board MED session protocols. At the same time, standard session protocols should still be easily readable for therapists and patients. No content should appear in session protocols that is potentially offensive or unforeseeable ("out of control") for patients.

A corresponding protocol function has been implemented in Tele-Board MED. The system can now generate protocols which include methods used, process phases and most important work results automatically when therapists follow some simple conventions in their work routines. As they use digital whiteboards cooperatively with patients, therapists can place a sticky note with a short description of the therapeutic method being used in the upper left corner of each page (or, rather, each whiteboard panel). For instance, such a sticky note can say "autogenic training" or "distinguishing between functional and dysfunctional beliefs". Many such sticky notes are already prepared as part of readily available therapy templates (see Sect. 3). Furthermore, in the lower right corner of each page (or panel) an icon can be placed to visualize the therapeutic method or the process phase of the treatment. Finally, therapist and patient can mark sticky notes or pictures as particularly important to include them straightforwardly in session protocols.

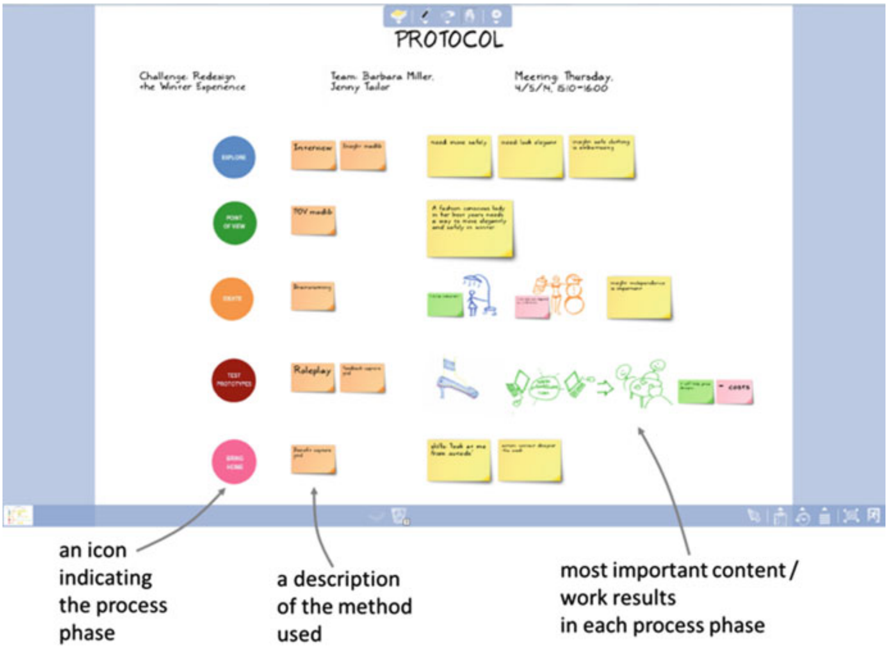


Fig. 3 Figure illustrates the revised format for automatic protocols, showing an example from one particular therapy session. The protocol indicates therapeutic process phases, methods used and most important session content/work results

On such a basis, Tele-Board MED generates protocols automatically which outline methods used, process phases and the most important therapy content of a treatment session (see Fig. 3 for an example).

Apart from that, an extra protocol template has been prepared for therapists to be filled out after therapy sessions. This template focuses on problems that a therapist has had. Resulting protocols are not meant to be read by patients. They are intended to support therapists who prepare for supervision (Fig. 4).

2.3 *Realizing an Automatic Protocol Function*

When doctors work with analogue patient files, these files typically contain sheets of paper with notes. Tele-Board MED stores patient records digitally. They contain whiteboard panels (corresponding to paper sheets) with notes, sketches and images.

In general, therapists can of course oversee the files of all patients they treat. However, during a therapy session, while a patient is being treated, only the file of this particular patient should be visible. The patient should not see files of other patients. Therefore, a new *treatment session mode* has been implemented in Tele-Board MED. At the beginning of each session, the therapist selects the pertinent

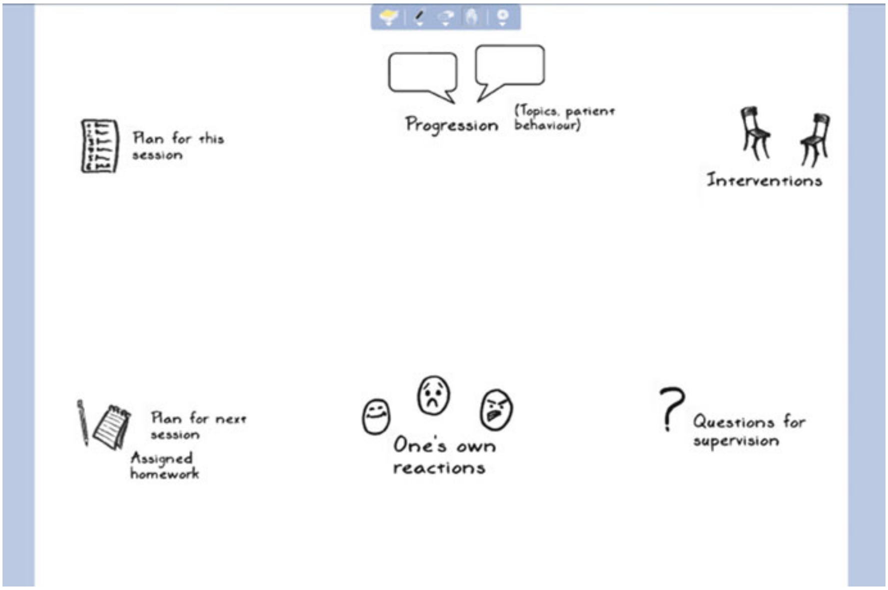


Fig. 4 A protocol template prepared for therapists to be filled out after treatment sessions—in particular when unexpected problems have occurred. The information gathered here is not intended for sharing with patients. Rather it should help in the preparation for supervisions

patient file, presses a "start session" button and all other patient files are hidden. At the end of each session, the therapist presses the "end session" button and can now access other patient files again.

Such a treatment mode is also the basis for the new protocol function: Each protocol provides an outline of all panels which were opened in a treatment session, i.e. while the treatment mode was on.

Figure 5 shows an example of how therapists can use the treatment session mode and create protocols subsequently.

In the protocol, all panels that have been opened during a session are listed chronologically. The panel that was opened first is listed first. Each panel (work sheet) can then be described and summarised in a number of ways. Users can decide whether or not to include a picture of the panel. Building on the features of the Tele-Board whiteboard client (Gericke and Meinel 2012; Gumienny et al. 2011), pinned or unpinned sticky notes can be included or excluded. Sticky notes can also be included when they have been marked as "important" by therapist or patient.

Furthermore, general information on the session can be included. Once again, users can adjust protocols according to their own preferences. For example, therapists can decide whether or not their protocols should include a reference to the patient, start and end times of therapies, panel names (corresponding to the headlines of paper sheets in analogue documentation) and so forth.

Technically, a protocol is a normal panel. Therefore, the automatically created protocol is not carved in stone. Its content can be modified easily by the user.

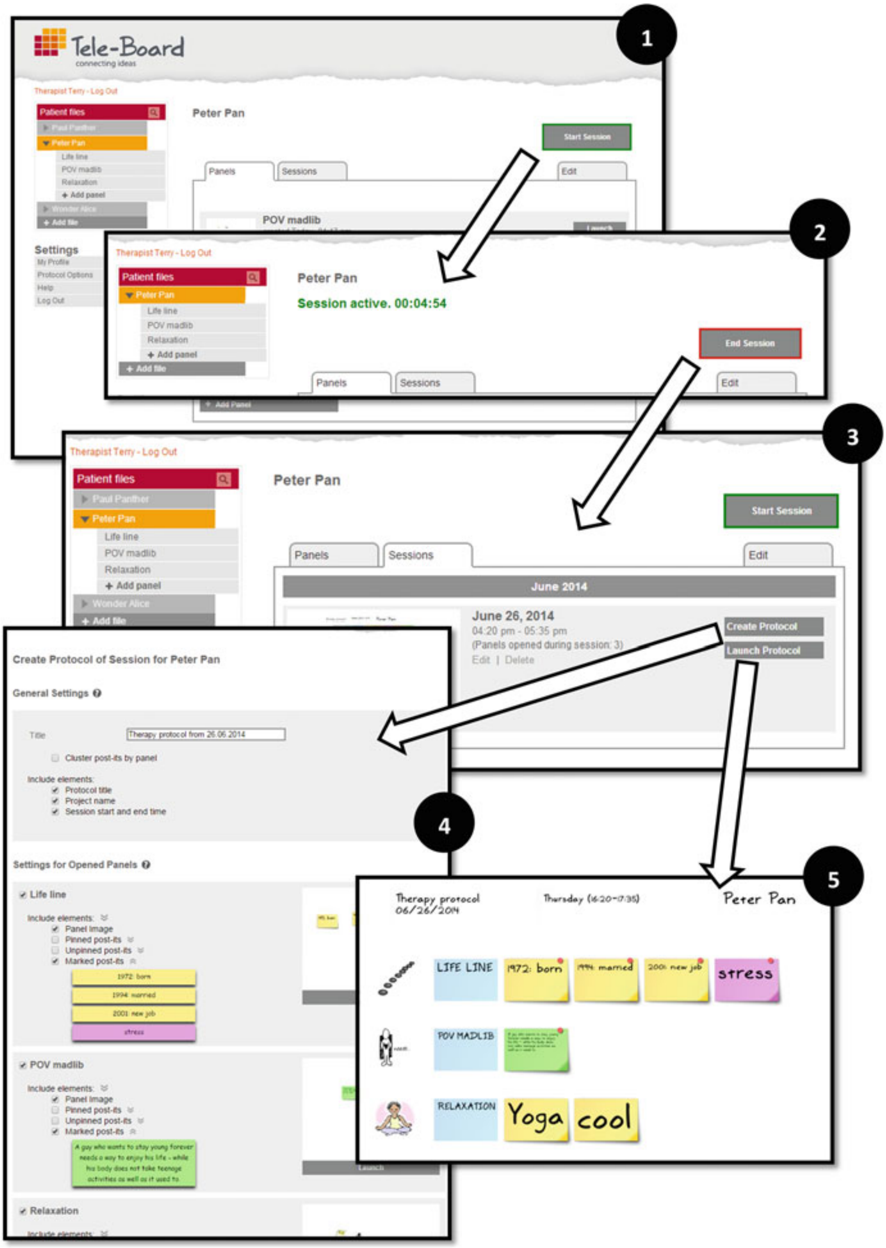


Fig. 5 Navigation through the user interface to create a protocol: (1) The therapist clicks on "Start Session". A clock starts running, indicating the length of the session. (2) Throughout the session, therapist and patient can work on one or more panels. In the example shown here, three panels are opened (however, the opened panels themselves are not shown in this figure). When the "End Session" button is pressed, a list of all sessions can be viewed that the therapist has had so far with this patient. (3) The therapist can select one session to open the corresponding protocol. She can either choose to see all options of automatic protocol creation again (4); in this case, she also sees a preview of how the protocol will look like depending on the options that she picks. Or (5) she can

2.4 *Building on a "Sandwich Metaphor": Suggestions for Using the Automatic Protocol*

In some ways, psychotherapeutic treatment sessions are meetings like many others: They start with initial hellos and introductory conversation, giving way to the main part of talking and eventually end with closing statements and saying goodbye.

There is a metaphor for this structure of typical conversations (maybe more prominently known as a metaphor in text writing); which is the "sandwich metaphor". We believe this metaphor can also help to describe ways in which Tele-Board MED protocols support psychotherapists: There is a main conversation part ("roast piece") which is embedded in intro- and outro-conversation ("two buns").

Almost anything can make a tasty sandwich as long as it is wrapped well in a bottom and a top bun. For therapy sessions, this means that a therapist will not launch a behaviour analysis before saying hello first. Rather, a therapist will start by welcoming the patient, asking how he or she is doing. A connection can be made to the last session; an overview of the current session can be given and so forth to create a pleasant intro. Correspondingly, treatment sessions should rather not end abruptly by simply dismissing the patient when time is over. Rather, summing up the session, agreeing upon some homework, discussing what one found particularly important etc. can help to create a smooth outro.

Tele-Board MED protocols should help prepare salubrious sandwiches. By building on the last sessions' protocol, therapists can easily recap where the team left off when starting a new session. Then, therapist and patient can document their new conversation digitally with little effort. In the end, automatic protocols can be a tool to close sessions smoothly. Instead of just sending the patient home, therapist and patient can look at the protocol jointly, changing it if necessary. The question "what should be in the protocol?" can serve as a prompt to quickly recapitulate the session. Therapist and patient can shortly reflect on issues they found most important. Afterwards, the therapist might even print out the resulting session protocol for the patient if he or she wants to take it home.

2.5 *Testing the Automatic Tele-Board MED Protocol Function*

To test the new Tele-Board MED protocol function, we introduced it to ten therapists in a 2-h timeslot. The therapists also tried out the system and its functionality themselves as part of the demo. Generally, the creation and use of

---

Fig. 5 (continued) start the protocol immediately—based on default options or previously saved settings. The protocol opens up as a new whiteboard panel automatic protocols was introduced in line with the “sandwich metaphor” (see Sect. 2.4) and several therapeutic use cases were demonstrated (see Sect. 3).

The sample of therapists who took part in this study resembles the sample from our prior test in 2013 (von Thienen et al. 2014a). Two therapists took part in both studies.

One aspect of the sample that is particularly important for the interpretation of results is the attitude towards technology. It is known that people who are technology-enthusiastic generally assign higher utility ratings to digital tools such as Tele-Board compared to people who are averse to technology (Chuttur 2009). In our sample from 2013, the average attitude towards technology was slightly positive: 0.68 in concrete numbers on a scale from −2 (hostile towards technology) to 2 (enthusiastic about technology). Using the same scale in 2014, the average attitude was slightly positive once again: 0.7 in concrete numbers, thus almost identical to the sample from 2013.

The average age was somewhat higher in the new study (38.2 years in 2014 versus 35.4 years in 2013). Correspondingly, the new sample group also has longer work experience. In 2014, 40 % of the therapists had a work experience longer than 2 years—compared to 27 % in 2013.

Regarding their present documentation habits and tools, most therapists in our study state they take handwritten notes while treating patients in sessions. The majority (60 %) writes an average of between ¼ and ½ of a DIN A4 page per session.

30 % of the therapists state they usually don’t add any more notes after sessions. 40 % typically add notes of about one quarter of a page, 30 % usually add one half of a page or more.

For the time being, both during sessions and afterwards handwriting is the most common means of documentation. 70 % of the therapists don’t ever use digital tools for their own case documentation either in or after treatment sessions.

Taking notes in treatment sessions typically leaves therapists with a somewhat unstructured session documentation. Most therapists go over their notes once again and create structured session protocols at least for some treatment sessions (see Fig. 6).

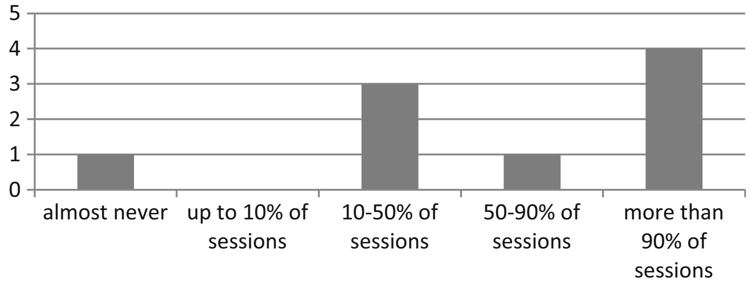


Fig. 6 Therapists answer the question “Have you created ‘orderly’ session protocols in the past, in addition to ‘spontaneous’ notes taken in or after sessions?”

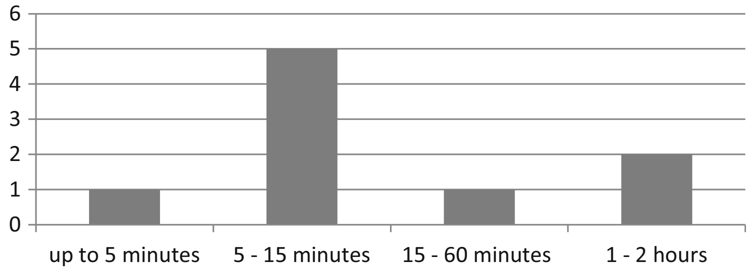


Fig. 7 Therapists answer the question “How much time do you typically need to write one session protocol?”

Table 1 Therapists explain why they write session protocols

Purpose	Average on a scale from 0 to 3	Standard deviation
To adhere to the laws	2.44	1.13
To get an overview of the treatment as a whole	2.22	0.67
To remember the case for the next meeting	2.1	1.1
To follow up on a session (What was important? What went well, what went badly?)	2.0	0.87
To prepare for supervision	1.33	1.12
To “round off” conversations with the patient: jointly summarize important issues and draw a conclusion	1.11	1.27
To hand out documentation to the patient	0.78	0.97

The scale ranges from 0 (not so important) to 3 (very important)

Figure 7 shows how much time therapists typically spend writing one session protocol. For the therapists in our sample, one therapeutic treatment session takes 50 min. This means, documenting one therapy session by writing a protocol takes about half as long as the therapy session itself. Thus, it is a considerable investment of time on behalf of the therapists.

We asked the therapists why they write protocols, suggesting the options named in Table 1. There was also the opportunity to add further purposes in free text fields; however no further purposes were added. Table 1 shows the hierarchy of reasons for writing session protocols as specified by the therapists.

Table 1 shows that adhering to legal documentation demands is quite important to the therapists; it is the highest ranked purpose. Handing session documentation out to patients—as though part of new legal demands—is the least important purpose for writing session protocols. Furthermore, joint documentation with the patient figures at the second lowest position.

In line with these findings, 70 % of the therapists state they have never handed out session protocols to patients. The other therapists state they have given out protocols to patients already, but just in very few cases. Against this background, the Tele-Board MED protocol function seems to make a great difference.

To start off, all therapists think they would work with session protocols regularly when supported by the automatic Tele-Board MED protocol function (Fig. 8). This makes a considerable difference given that so far, without Tele-Board MED, some therapists don't work with session protocols on a regular basis.

Furthermore, therapists are likely to share their session protocols with patients much more frequently when using Tele-Board MED (see Fig. 9).

One issue that is often sensitive for automated functionality is the quality of results. For instance, the program MS Word offers automatic summaries of text documents. However, these automatic summaries are often cumbersome. They don't meet the quality of human-written text summaries.

In Tele-Board MED, there are some minute actions that therapists can take to prepare valuable automatic protocols. For instance, it helps to give comprehensible

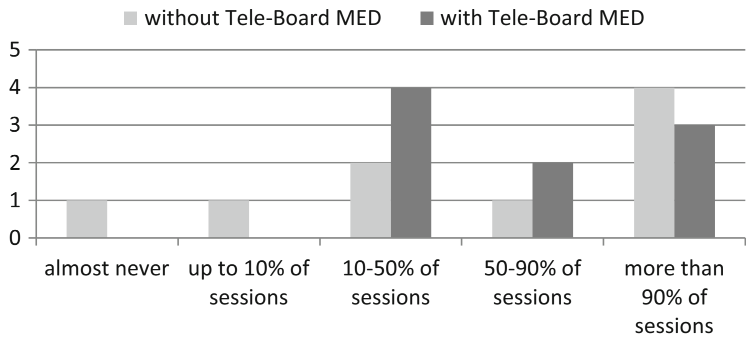


Fig. 8 The *light bars* show how often therapists write session protocols nowadays. The *dark bars* show estimates of the therapists: how often they would work with session protocols when supported by the automatic Tele-Board MED protocol function

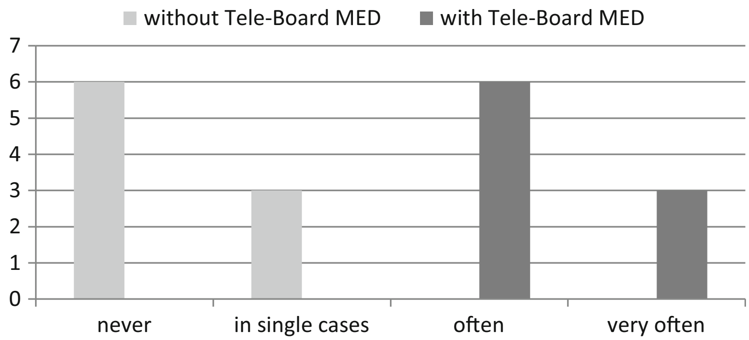


Fig. 9 Before Tele-Board MED had been introduced, therapists were not in the habit of sharing session protocols with patients. The *light bars* show answers to the question: "Have you ever handed out session protocols to patients?" With Tele-Board MED, therapists are likely to hand out session protocols to patients. The *dark bars* show answers to the question: "Do you think you would hand out Tele-Board MED session protocols to patients?"

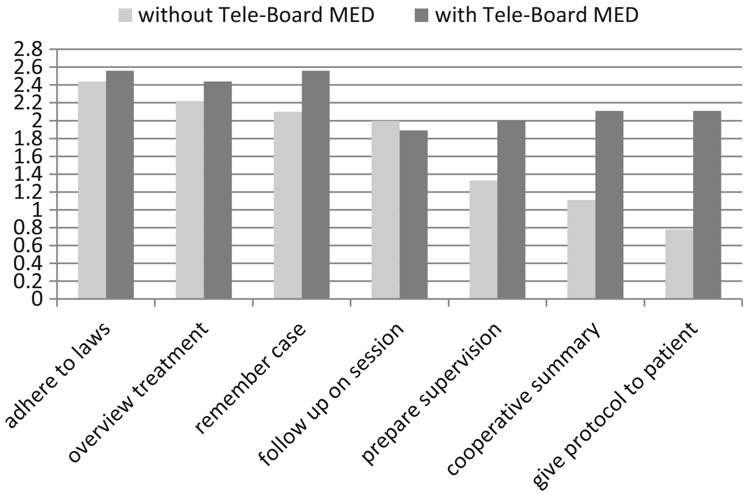


Fig. 10 Without Tele-Board MED, therapists use manually-created session protocols for a limited number of purposes. The *light bars* show answers to the question: “For what purposes do you create session protocols?” Answers range from 0 (not so important) to 3 (very important). By contrast, *dark bars* show answers to the question: “For what purposes would you use automatically created Tele-Board MED session protocols?” Once again, answers range from 0 (not so important) to 3 (very important)

names to new panels or to work with prepared templates. Such minute actions had been included in the demo of our feedback study. Thus, easy to read protocols were obtained even though the content was not pre-planned; the therapists could enter information spontaneously.

We asked the therapists for the ways in which they would want to use Tele-Board MED protocols. It was striking that they found more reasons to use such automatic protocols than they had seen reasons for self-created protocols (Fig. 10). Thus, automatic Tele-Board MED protocols seem to be at least as serviceable as self-written protocols: They seem to serve even more purposes.

Remarkably, the therapists show strong consensus when discussing reasons for using Tele-Board MED protocols. The highest standard deviation among all seven purposes (as shown in Fig. 10) is 0.78. By contrast, the therapists seem to have rather dissimilar reasons for creating session protocols manually. As shown in Table 1, standard deviations often exceed 1.0 when therapists report their reasons for (manually) writing protocols.

With Tele-Board MED, handing out session protocols to patients or summarising session content cooperatively with the patient become much more recognized purposes (cf. Fig. 10). The average rating of importance for handing out session protocols to patients was 0.78 without Tele-Board MED; the average rating of importance for summarising session content cooperatively with the patient was 1.1. Both numbers move up to 2.11 with Tele-Board MED, on a scale of 0 (not so important) to 3 (very important).

Taking into account qualitative feedback as well, we conclude that the protocol function is strongly appreciated by the therapists. It allows them to create official session protocols much more frequently than before—given that writing protocols has been a time-consuming task up to the present. Furthermore, automatic Tele-Board MED protocols can serve at least as many purposes as present-day, human-written protocols. Finally, Tele-Board MED helps to meet important goals of patient empowerment initiatives or patient rights laws. Therapists say they would share treatment protocols with patients when supported by Tele-Board MED. Thus, documentation can become much more transparent and cooperative with the new technical possibilities.

3 Tele-Board MED Supports Documentation in Diverse Behaviour Psychotherapeutic Settings

Tele-Board MED seems well equipped to document therapeutic interactions that can take place directly at a digital whiteboard. By means of the history browser users can scan all their prior digital activities. The automatic protocol function summarises content that has been entered into Tele-Board MED during a therapy session. However, there can be many psychotherapeutic situations when therapist and patient spend their time away from the whiteboard. That might be the case when a patient practices autogenic training, to give an example. Furthermore, some important documentation activities take place after therapy sessions—such as the writing of case reports for insurance companies.

This section describes five standard psychotherapy settings. In some settings, patient and therapist might gather at a digital whiteboard, in other settings they would typically put the whiteboard aside. Some settings occur in therapy sessions, others occur afterwards. Tele-Board MED supports documentation in all these use cases.

3.1 Collecting Information for Case Reports

To pay for treatments, health insurance companies typically demand elaborate case reports. These reports should clarify whether the patient does indeed have a recognized mental illness, which is usually considered the case if he or she fulfils a sufficient number of diagnostic criteria specified either in the *International Statistical Classification of Diseases and Related Health Problems* (ICD) by the World Health Organization (e.g., 1992) or the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) of the American Psychiatric Association (e.g., 2013). Furthermore, the proposed treatment needs to be promising; its efficacy should be scientifically confirmed.

In Germany, concise requirements have been issued by the Kassenärztliche Bundesvereinigung (1992). These directives are valid for all statutory insurance companies. All behaviour psychotherapists need to address the following issues when they want insurance companies to pay for more than 25 treatment sessions (Müther 2009):

1. Symptoms described by the patient
2. Case history of the patient/biographical anamnesis
3. Statement regarding mental state
4. Statement regarding somatic state
5. Behaviour analysis including a SORKC analysis
6. Diagnosis
7. Therapy goals and prognosis
8. Treatment plan
9. Justification of a prolonged treatment

To support therapists who need to write case reports at some point, we have devised templates that help gather the necessary information systematically. For all issues that need to be addressed in behaviour psychotherapeutic case reports, there are multiple templates to select from (see Fig. 11). *A therapist, who uses one template of each group over time, will have all the necessary information available in a digital format when he or she wants to assemble a case report.*

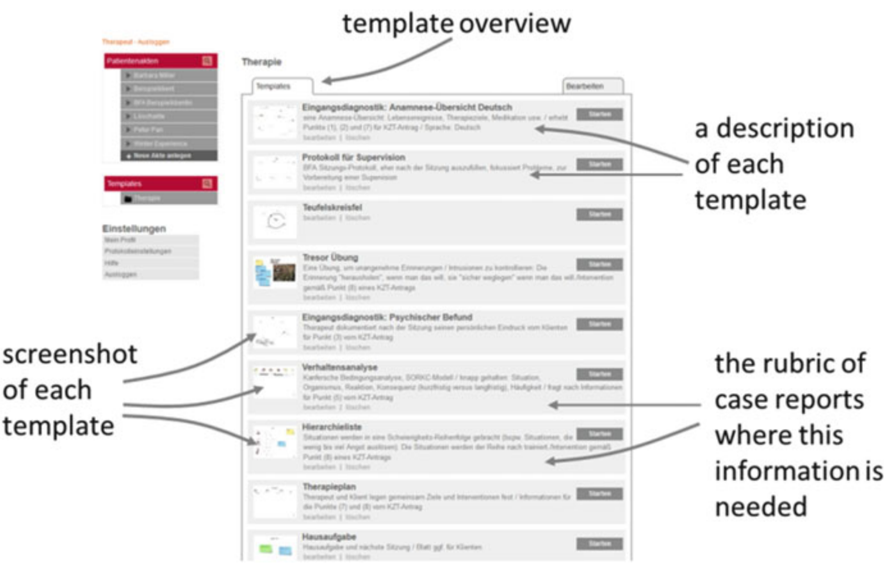


Fig. 11 The template overview shows a screenshot of each template. A verbal description explains what each template is typically used for. Templates which help to gather or analyse information are categorised according to the rubrics of case reports, when such information is requested

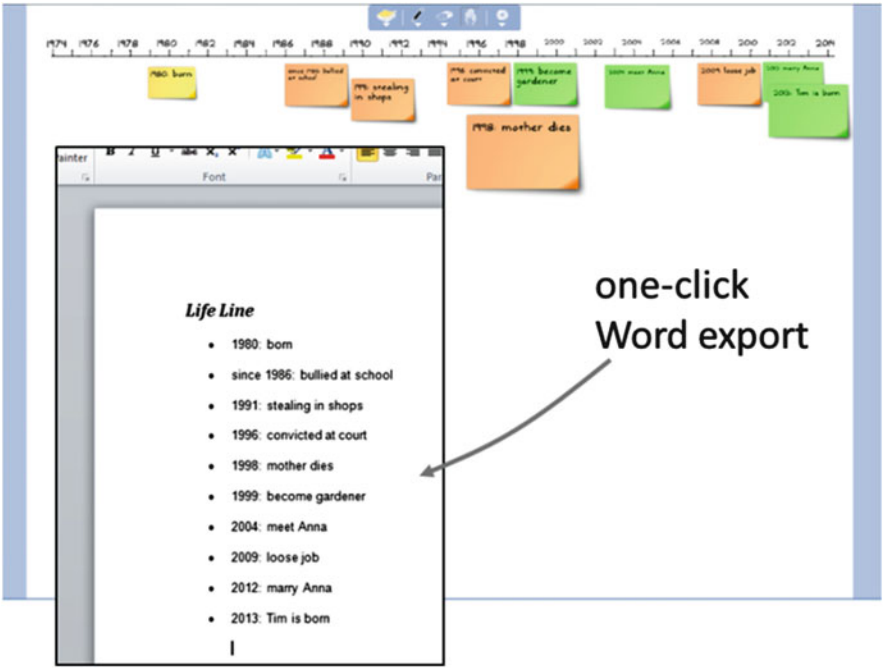


Fig. 12 Tele-Board MED makes it easy to gather and sort information. Using the Tele-Board MED Word Export function, all information can be transferred to a text file immediately

Using an MS Word export function, information that has been gathered in Tele-Board MED can be exported to Word easily. Each Word file contains headlines followed by lists of keywords or short statements (Fig. 12). Thus, to obtain a complete case report, therapists simply needs to cast the keywords into a running text.

Compared to documentation procedures formerly used in psychotherapy, there is (1) no need to search for information in long handwritten files, (2) no need to digitalize handwritten information, (3) no surprising discovery that one forgot to ask about some pieces of information that are required in a case report.

While the templates provided by us are tailored to support behaviour psychotherapies, additional templates can be created easily. Thus, psychotherapists who follow approaches other than behaviour psychotherapy, or medical practitioners of basically any field, can easily assemble templates which ease their work as well. The provided templates can also be redesigned, i.e. therapists can easily adjust, amend or change them completely according to their own preferences.

3.2 *Using Analysis Schemes*

Behaviour psychotherapy is well equipped with analysis schemes that have proven helpful in decades of psychotherapy practice and research. The SORKC model (Kanfer and Saslow 1965), the ABC analysis (Ellis 1973) or plan analysis (Caspar 1986) are just some examples. Tele-Board MED provides several templates which support popular analyses. For instance, there are short and long versions of the SORKC model. There are also different types of examples to account for differing preferences and thinking styles of patients such as examples with animals as opposed to examples with machines (see Fig. 13).

Apart from behaviour psychotherapy, design thinkers also use many analysis schemes that have proven helpful in years of practice. We have adapted several design thinking based analysis schemes for psychotherapy sessions. In particular, there are templates which support processes of problem solving (von Thienen and Meinel 2014, in press).

Tele-Board MED analysis templates help to address four needs at once. First, the templates help to introduce, explain and carry out an analysis. Second, the analysis result is perfectly documented without any extra effort: All content is stored automatically. It can be opened up and elaborated on in forthcoming sessions. Third, a session protocol is created automatically which states that this analysis has been carried out in that session. Fourth, each analysis can easily be sent to a Word file and included in a case report.

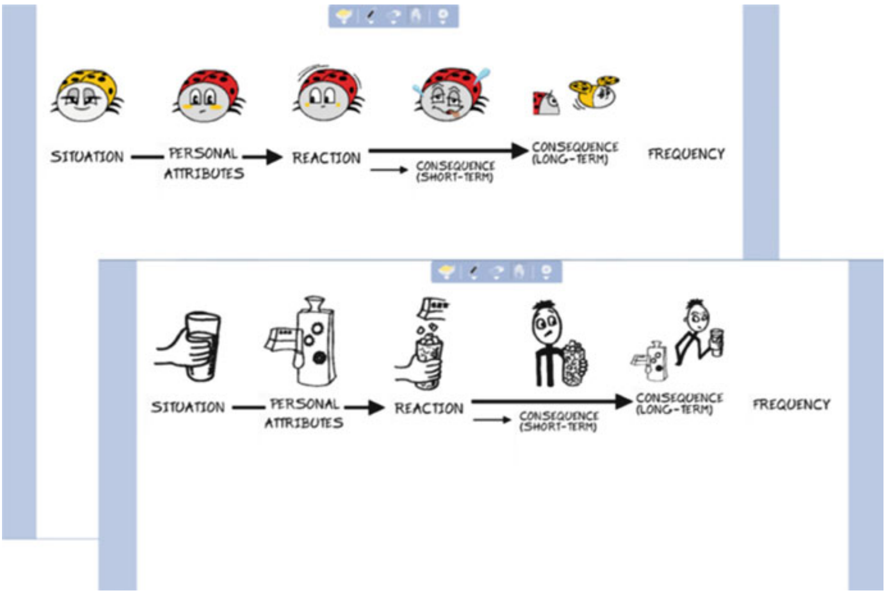


Fig. 13 To explain the SORKC model to patients with differing thinking styles, the therapist can choose between several analysis templates. Here, two templates for short SORKC analyses are shown, with animal or machine examples

3.3 Supporting Exercises

Many psychotherapeutic treatments go beyond talking. People also engage actively in psychotherapeutic exercises. For instance, patients can be invited to practice some relaxation technique. In other cases, patients can practice how to utter criticism in a constructive way. Or, patients learn how to approach a dreaded object step by step. There are standard exercises that have been found effective in research and many therapists use them with only minor adaptations. We provide templates that support several of these standard exercises (see Fig. 14 for an example). In addition, therapists can easily adjust the templates to suit their own preferences. They can also add templates for other exercises.

Such exercise templates can support therapeutic treatments in multiple ways. A therapist can open an exercise template in a session. Even if no one ever looks at the template throughout the session (e.g., because the monitor is switched off), a session protocol is created automatically that states the exercise has been used in that session. Thus, with only one click there is a lot of sensible session documentation.

Some exercises come with long texts to be spoken by the therapist. Therapists can read the text from the templates if they wish. When patients close their eyes (e.g., in relaxation exercises), they might not even notice text is read to them and not memorized by the therapist.

Therapists can open exercise templates to explain to the patient what is going to happen next. In case a therapist is rather unfamiliar with an exercise herself, she can use the template not only to inform the patient about what is going to come next, but also to recall the details herself. At the end of a session, the template can be printed out to support a patient who should then practice the exercise at home.

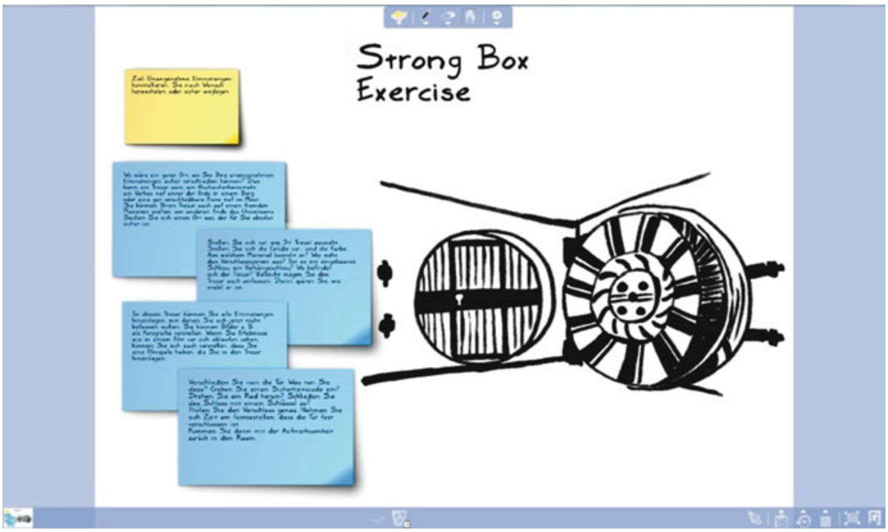


Fig. 14 A sample exercise template. It includes a brief tutorial for novices (four *blue/dark* instruction cards on the *left*) plus a short methodological description of the procedure and its goals (*yellow/light* method card in the *upper left corner*) (Color figure online)

3.4 Supporting Psycho-Education

Providing background information on some subject is a common measure in behaviour psychotherapy. A patient who suffers from anxiety can be introduced to the vicious cycle of anxiety (e. g., Margraf and Schneider 1990). A person who suffers from burnout can be introduced to one or the other stress model (e.g., Lazarus 1999). A drug addict can be introduced to models of health behaviour change (e.g., Prochaska and Velicer 1997). Tele-Board MED provides several templates that support psycho-education (see Fig. 15 for an example). They can also be printed out for the patient to take them home after a session.

Such templates support psychotherapists in four important ways. First, they help to explain the subject of psycho-education with easy to understand examples. Second, they can help a patient who needs to relate the provided information to his own case. There is space left blank for the patient to note down what is the case in his life. Third, a session protocol is created automatically which documents that psycho-education on this particular subject was part of the session. Fourth, when the patient has entered information about his own case (e.g. what a typical vicious cycle looks like in his life), this information can easily be exported to Word for use in a case report.

As always, templates can be easily changed by the therapist and new templates can be added easily.

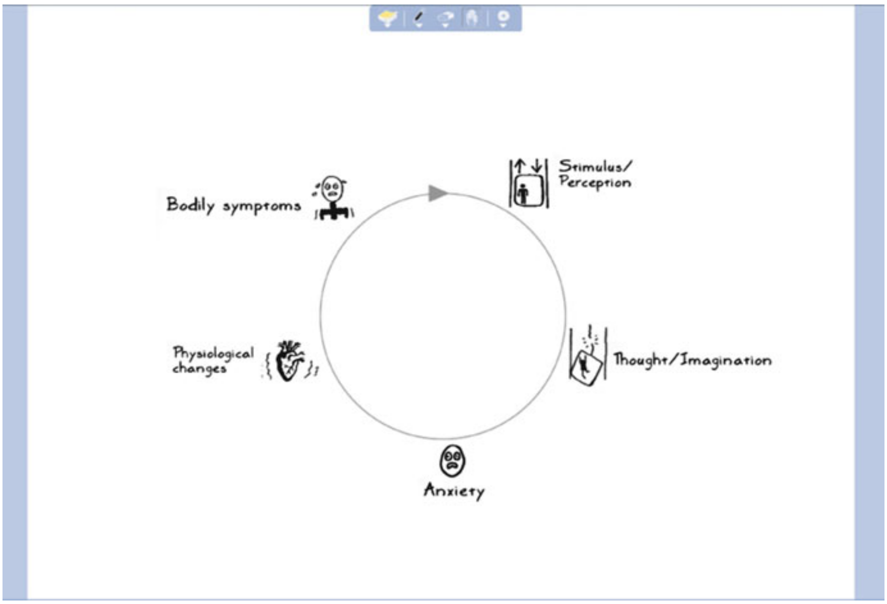


Fig. 15 A sample template for psycho-education. Here, a vicious cycle of anxiety is described

3.5 Sessions with a Lot of Interaction Apart from Tele-Board MED

Many sessions do not primarily focus on gathering information. Rather, it is important to create rich experiences for the patient. In one case, a patient might cast out his life line, reflecting on his burdening and uplifting experiences (Schauer et al. 2011). Or, a patient and his therapist arrange two chairs in the room. One chair is said to belong to the patient's father. The other chair is said to belong to the child that the patient used to be when he was little. Then, therapist and patient switch back and forth between the chairs and stage a dialogue between the two protagonists. Thus, they try to learn about unmet needs that the patient had when he was little (Young et al. 2013).

Commonly, therapist and patient spend such sessions away from the whiteboard. Nonetheless, Tele-Board MED can help to document these sessions. There is a handy feature to integrate analogue work in digital records: the Tele-Board sticky pad application. This application for mobile devices does not only allow the user to send digital sticky notes to the whiteboard panel, but also pictures. For a minimum of documentation, the therapist simply needs to take a picture of the setting with a mobile device such as a tablet computer or a smartphone. With one click, it can be sent to the whiteboard. Adding a comprehensible panel name such as "life line" will suffice to create a session protocol that captures crucial information (see Fig. 16).

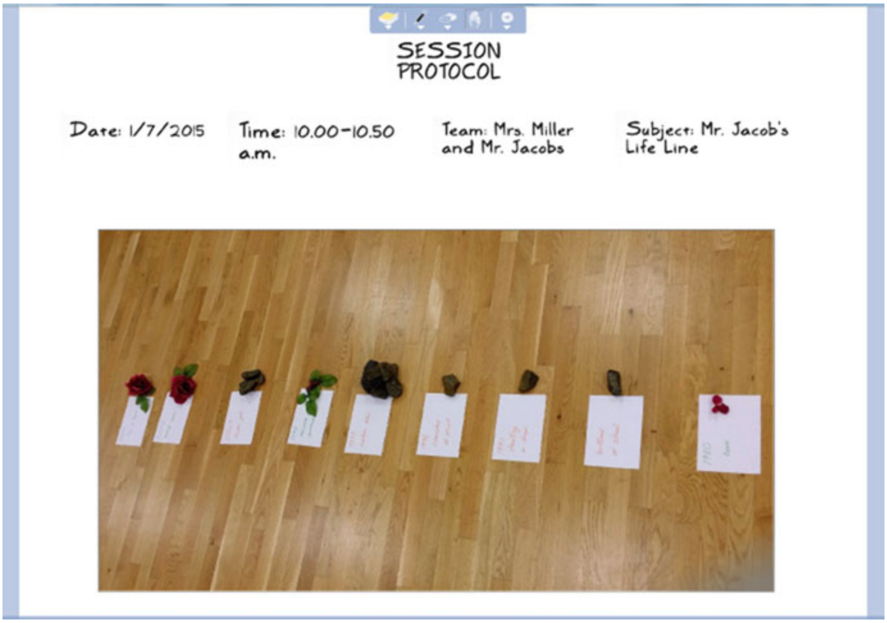


Fig. 16 Session protocols such as the example shown here can be created in a few seconds. Such a protocol includes the date of the session, start and end time, name of therapist and patient (or pseudonyms, if preferred), the session subject (such as "life line") and a picture showing work results—here the depicted life line of the patient

In other cases, therapist and patient might use the last 5 min of a session to record the insights they found most striking.

4 Contributions to Design Thinking Documentation

So far, documentation solutions have been sketched out which support behaviour psychotherapists. However, this chapter will close by returning to the field where Tele-Board (MED) originates from: design thinking. After all, new features that have been developed for Tele-Board (MED) can be used by design thinkers too.

Documentation needs are manifold in the design thinking community and so is the research regarding new documentation solutions. We will quickly survey several design thinking projects on documentation and discuss how Tele-Board (MED) relates to some of their central concerns. In particular, we will outline how the newly developed automatic protocol function can help to address important needs that are intensively investigated by the design thinking community at present.

Retrospect understanding Gericke et al. (2012, 2014) want to support design thinkers and externals who wish to understand creative sessions and their outcomes in retrospect. This should be achieved by saving enough information *while teams are working* so that questions which arise later on can be answered on the basis of readily available information. By way of contrast, it should not be necessary to contact design thinkers and interview them to understand past decisions or actions in their design projects.

To accomplish this documentation challenge, Gericke et al. (2012, 2014) have developed the Tele-Board history browser (as described in Sect. 1.1). Tele-Board stores all actions that teams carry out on digital whiteboards. The history browser helps to review these activities such as adding sticky notes, moving them around or deleting them. Furthermore, additional synchronized collaboration data, such as videos and information regarding the audio level can help to identify or recall important moments in the work process.

Finding important information artefacts and tracing project paths In the course of design thinking projects, teams typically accumulate huge amounts of information artefacts such as sticky notes, whiteboard drawings, pictures, paper prototypes etc. These huge amounts of information can barely be taken into account all at once. Both, design teams or externals who try to understand the path of project progression can experience an information overload.

To counter such an information overload, Voget (2013) suggests *ConnectingInfos*—a software that stores and organizes information artefacts. “It provides a project overview, where all acquired information is displayed on a time line. ConnectingInfos supports the ability of seeing a big picture by highlighting important information and displaying associations between information items. Users can avoid the danger of information overload by reducing the amount of information displayed [. . .]. Using logistic regression analysis, we predict the importance of information items using a variety of predictor variables, such as age of the information item or number of keywords in the title" (p. iii).

Knowledge handover In design thinking projects, there often needs to be a transfer of knowledge from the design team to externals such as companies that might turn prototypes into widely available products. Beyhl et al. (2013) investigate how the information handover from design thinking teams to engineers can be improved. "At the end of a design thinking project, the final outcome of the project is often presented to the external client, who decides about the realization of the product or service. This presentation includes the final prototype, which fulfills the purpose of illustrating the final idea. Therefore, the final presentation passes on the overall idea, but often neglects design rationales [. . .] [;] engineers engaged by the client make their decisions based on incomplete knowledge when realizing the product or service. [. . ..] The engineers may end up creating a less desirable product or service" (p. 1).

To help design thinking teams pass over more knowledge without much extra documentation efforts, Beyhl and Giese (2014) suggest a software solution. Their "documentation platform consists of a graphical user interface (GUI), an aggregator and an active repository. The GUI enables Design Thinkers to organize their artifacts (e.g. photographs or post-it clusters) captured by the aggregator, which loads these artifacts from arbitrary software tools (e.g. online storage services) used by Design Thinkers. The active repository is a storage whose content is steadily analysed by an inference engine, which extracts knowledge embodied within the stored artifacts" (p. 214). For instance, pictures of sticky notes might be entered into the system. The system might extract the knowledge that a certain cluster of sticky notes belongs to one particular process phase. Furthermore, when the cluster shows up in more than one process phase, it can be assumed that an information hand-over has occurred from one phase to the next.

Revealing design rationales and increasing mindfulness According to Menning et al. (2014), design thinking documentation should serve three purposes. First, the design rational should become apparent; there should be an explicit documentation of the reasons for particular design decisions. Second, documentation should support reflection on what went well or less well in the design project and why. In addition, teams should reflect on changes in their supposed design mission (reframing). Third, documentation should help design teams communicate the reasons for their design choices and design solutions to externals.

To support sensible design documentation, Menning et al. (2014) have created a protocol template to be filled out by design teams on a regular basis—such as daily. It is typically made available on DIN A3 paper sheets. The template asks teams to document their point of departure and what they did; teams should evaluate their findings, plan next steps and reflect upon their design path.

Visualising process phases, methods used and most important work results In a project which is just starting, Giese (2014) and his team investigate "how to capture and visualize employed DT methodologies at work to understand them in detail. For example, this includes capturing employed methodology phases, methods and techniques, as well as artifacts in which the outcome of certain activities is manifested."[1] Since this project has only just begun, no concrete documentation solution is suggested yet.

The new Tele-Board (MED) protocol function addresses several needs that are currently tackled in design thinking research. In terms of the output, our approach resembles that of Giese (2014) and his team, as well as that of Voget (2013). Automated Tele-Board (MED) protocols visualize process phases, methods used and most important work results—such as sticky notes with insights or pictures of prototypes (Figs. 3 and 5 showed sample protocols).

Furthermore, many design thinkers assume that documentation should demand as little extra work of teams as possible. Otherwise, documentation might not happen at all. For instance, with respect to educational design thinking settings Beyhl et al. (2013) remark: "The biggest challenge [...] is to get students to document voluntarily" (p. 4). That is one of the reasons why the authors find that a workable documentation solution "has to be integrated into the overall process unobtrusively" (p. 4). For similar reasons, the documentation solution of Gericke et al. (2012, 2014) does not demand any extra documentation activity of design thinkers at all.

The solution suggested by Menning et al. (2014) takes a slightly different turn. Their approach does call for a considerable amount of documentation effort on behalf of the teams. Protocols have to be filled out manually on a regular basis. However, this approach envisions immediate benefits for the documenters. They become more mindful of the process, of the reasons for design decisions, of expanding knowledge or competencies.

Depending on user preferences, Tele-Board (MED) protocols can be both completely unobtrusive and labour-free or purposefully applied documentation tools which support reflective team activities. To obtain labour-free protocols, teams can simply carry out their project work on the digital whiteboards of Tele-Board. Protocols of work sessions are then created and stored automatically. However, teams can also work actively with the protocols, e.g. amend or refine them. Thus, protocols can support reflective team activities in line with the design thinking motto "be mindful of [the] process" (d.school 2011, p. 0). Accordingly, we have suggested a usage of Tele-Board (MED) protocols building on the "sandwich metaphor" in Sect. 2.4: Protocols can help recall earlier works, document current activities and they can support reflections at the end of a session.

With some minor adjustments, Tele-Board (MED) protocols could help to address even more needs. For instance, design teams could be cued to reflect on

[1] https://hpi.de/de/dtrp/projekte/projekte-201415/the-design-thinking-methodology-at-work-capturing-and-understanding-the-interplay-of-methods-and-techniques.html

important design decisions by making this a standard part of session protocols. At the end of each design thinking day, teams could look at their automatic protocols asking themselves whether the reasons for their design choices are documented already. Elsewise teams could simply add one or two sticky notes on the subject. Furthermore, protocols could be optimized for information handover to externals or to understand projects in retrospect.

Generally, Tele-Board (MED) has the advantage of supporting design work proper and design documentation at once. Therefore, teams are not confronted with additional tools such as paper based design diaries or software solutions for documentation alone, which might be put aside for reasons of seeming convenience. Thus, teams can use Tele-Board (MED) for reasons of design support. Documentation occurs automatically then, without extra tools. Users can decide for themselves how much to engage in additional documentation activities, e.g. to be more mindful of the process and to reflect on their design work.

In sum, the project of adapting Tele-Board for behaviour psychotherapy has not detached the system from documentation needs in design thinking. Rather, the system has acquired new features that can be serviceable for design thinkers too. Data can be saved more securely now (Perlich et al. in press). There is more support for information analyses and information export (cf. sects. 3.1 and 3.2). Finally, maybe most importantly, there is a protocol function that helps teams to document their project work in terms of process phases, methods used and central insights almost automatically.

References

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders. APA, Washington, DC

Aujoulat I, D'Hoore W, Deccache A (2006) Patient empowerment in theory and practice: polysemy or cacophony? Patient Educ Counsel 66(1):13–20

Beyhl T, Berg G, Giese H (2013) Towards documentation support for educational design thinking projects. In: Proceedings of the international conference on engineering and product design education, pp 408–413

Beyhl T, Giese H (2014) Connecting designing and engineering activities II. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Building innovators. Springer, Berlin, pp 211–239

Bundesärztekammer, Kassenärztliche Bundesvereinigung (2008a) Empfehlungen zur ärztlichen Schweigepflicht, Datenschutz und Datenverarbeitung in der Arztpraxis. Deutsches Ärzteblatt 105(19):A 1026–A 1030

Bundesärztekammer, Kassenärztliche Bundesvereinigung (2008b) Technische Anlage zu den Empfehlungen zur ärztlichen Schweigepflicht, Datenschutz und Datenverarbeitung in der Arztpraxis. Deutsches Ärzteblatt 105(19):1–12

Bundes-Patientenrechtegesetz (2013) Gesetz zur Verbesserung der Rechte von Patientinnen und Patienten. Bundesgesetzblatt Teil I Nr. 9 ausgegeben zu Bonn am 25. Februar 2013

Caspar F (1986) Die Plananalyse als Konzept und Methode. Verhaltensmodifikation 4:235–256

Chuttur MY (2009) Overview of the technology acceptance model: origins, developments and future directions. Sprouts: working papers on information systems, vol 9, no 37 d.school – Hasso Plattner Institute of Design at Stanford (2011) Bootcamp bootleg. http://dschool.stanford.edu/wp-ontent/uploads/2011/03/BootcampBootleg2010v2SLIM.pdf. Accessed 26 Nov 2014

Ellis A (1973) Humanistic psychotherapy. The rational-emotive approach. McGraw-Hill, New York

Gericke L, Gumienny R, Meinel C (2012) Tele-Board: following the traces of your design process history. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Studying co-creation in practice. Springer, Berlin, pp 15–29

Gericke L, Meinel C (2012) Understanding asynchronous design work—segmentation of digital whiteboard sessions. In: Proceedings of the IEEE 13th international conference on information reuse & integration (IRI), pp 700–705

Gericke L, Wenzel M, Meinel C (2014) Asynchronous understanding of creative sessions using archived collaboration artifacts. In: Proceedings of 2014 international conference on collaboration technologies and systems (CTS), pp 41–48

Giese H (2014) The design thinking methodology at work—capturing and understanding the interplay of methods and techniques. Abstract of a new project. https://hpi.de/de/dtrp/projekte/projekte-201415/the-design-thinking-methodology-at-work-capturing-and-understanding-the-interplay-of-methods-and-techniques.html. Accessed 26 Nov 2014

Gumienny R, Gericke L, Wenzel M, Meinel C (2012) Tele-Board in use: applying a digital whiteboard system in different situations and setups. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Measuring performance in context. Springer, Berlin, pp 109–125

Gumienny R, Meinel C, Gericke L, Quasthoff M, LoBue P, Willems C (2011) Tele-Board: enabling efficient collaboration in digital design spaces across time and distance. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Understand—improve—apply. Springer, Berlin, pp 147–164

Kanfer FH, Saslow GH (1965) Behavioral analysis: an alternative to diagnostic classification. Arch Gen Psychiatry 12:529–538

Kassenärztliche Bundesvereinigung (1992) Hinweise zum Bericht des Therapeuten und zur Begutachtung bei Anträgen auf Verhaltenstherapie. Deutsches Ärzteblatt 89(31/32):A1 2642–2643

Koch S (2012) Improving quality of life through eHealth—the patient perspective. In: Proceedings of the European medical informatics conference (MIE 2012). Quality of life through quality of information. Studies in health technology and informatics, vol 180, pp 25–29

Lazarus RS (1999) Stress and emotion. A new synthesis. Free Association, London

Leiner F, Gaus W, Haux R, Knaup-Gregori P, Pfeiffer K-P (2009) Medizinische Dokumentation. Lehrbuch und Leitfaden. Schattauer, Stuttgart

Margraf J, Schneider S (1990) Panik. Angstanfälle und ihre Behandlung. Springer, Berlin

Menning A, Beyhl T, Giese H, Weinberg U, Nicolai C (2014) Introducing the Logcal: template-based documentation support for educational design thinking projects. In: Proceedings of the 15th international conference on engineering and production design education, pp 148–154

Miller RH, Sim I (2004) Physicians' use of electronic medical records: barriers and solutions. Health Aff 23(2):116–126

Müther M (2009) Bericht an den (VT) Gutachter. Schneller, leichter, kompetent. Deutsche Gesellschaft für Verhaltenstherapie, Tübingen

Perlich A, von Thienen JPA, Meinel C (2014) Supporting talk-based mental health interventions with digital whiteboards. In: Proceedings of the European medical informatics conference (MIE2014), pp 433–437

Perlich A, Sapegin A, Meinel C (in press) Implementation of data security requirements in a web-based application for interactive medical documentation. In: Proceedings of the international conference on health informatics (HEALTHINF)

Plattner H, Meinel C, Weinberg U (2009) Design thinking. Mi-Wirtschaftsbuch, München

Prochaska JO, Velicer WF (1997) The transtheoretical model of health behavior change. Am J Health Promot 12(1):38–48

Schauer M, Neuner F, Elbert T (2011) Narrative exposure therapy (NET). A short-term intervention for traumatic stress disorders. Hogrefe & Huber, Cambridge

Voget L (2013) An Association-based documentation tool to support design thinking groups in using their acquired knowledge. Master thesis, Hasso Plattner Institute at the University of Potsdam, Germany

von Thienen JPA, Meinel C (2012) Using Tele-Board to bring (some) design thinking into clinics. Part of the series: Tele-Board MED—rethinking patient files. Presentation available online at http://www.tele-task.de/de/archive/lecture/overview/7949

von Thienen JPA, Meinel C (2014) A design thinking process to tackle individual life problems (created for use in behaviour psychotherapy). Electronic colloquium on design thinking research: http://ecdtr.hpi-web.de/report/2014/002

von Thienen JPA, Meinel C (in press) Tele-Board MED: supporting creative problem solving in behaviour psychotherapy. In: Corraza GE, Agnoli S (eds) Multidisciplinary contributions to the science of creative thinking. Springer

von Thienen JPA, Meinel C, Nicolai C (2014a) How design thinking tools help to solve wicked problems. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Building innovation eco-systems. Springer, Cham

von Thienen JPA, Perlich A, Meinel C (2014b) Tele-Board MED: supporting twenty-first-century medicine for mutual benefit. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research. Building innovators. Springer, Berlin

World Health Organization (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. WHO, Geneva

Young JE, Klosko JS, Weishaar ME (2013) Schema therapy. A practitioner's guide. Guilford, New York

Preserving Access to Previous System States in the Lively Kernel

Lauritz Thamsen, Bastian Steinert, and Robert Hirschfeld

Abstract In programming systems such as the Lively Kernel, programmers construct applications from objects. Dedicated tools make it possible to manipulate the state and behavior of objects at runtime. Programmers are encouraged to make changes directly and receive immediate feedback on their actions. However, when programmers make mistakes in such programming systems, they need to undo the effects of their actions. Programmers either have to edit objects manually or reload parts of their applications. Moreover, changes can spread across many objects. As a result, recovering previous states is often error-prone and time-consuming. This report presents an approach to object versioning for systems like the Lively Kernel. Access to previous versions of objects is preserved using version-aware references. These references can be resolved to multiple versions of objects and, thereby, allow reestablishing preserved states of the system. We present a design based on proxies and an implementation in JavaScript.

1 Introduction

Programming systems such as Squeak/Smalltalk (Ingalls et al. 1997; Goldberg and Robson 1983) and REPLs for LISP or Python allow adapting programs at runtime. Changes to programs in such environments are effective immediately and programmers can see or test right away what differences their actions make. Thus, these systems provide immediate feedback to programmers.

A subset of such systems, which includes, for example, Self (Ungar and Smith 1987, 2007) and the Lively Kernel (Ingalls et al. 2008; Krahn et al. 2009), are those built around prototype-based object-oriented languages (Lieberman 1986). In prototype-based systems programmers create applications using objects and without having to define classes first. In Self and the Lively Kernel, programmers can

L. Thamsen • B. Steinert • R. Hirschfeld (✉)
Software Architecture Group, Hasso Plattner Institute, University of Potsdam, 14482 Potsdam, Germany
e-mail: robert.hirschfeld@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_15

inspect and change the state and behavior of objects at runtime. Programmers create actual objects, not source code that only abstractly describes potential objects.

The Lively Kernel was designed to support this kind of development (Lincke and Hirschfeld 2013). It provides tools to directly manipulate the style, composition, and scripts of graphical objects. For example, programmers can change the positions and composition of objects directly using the mouse. They can use temporary workspaces to manipulate objects programmatically. They can edit and try methods directly in the context of graphical objects.

For example, to add new functionality to a graphical application, a Lively Kernel user might copy an existing button object and then modify the new button object. This could entail moving the new button to a sensible position, resizing it, setting a new label, and adding a script to be executed on mouse clicks. The user makes all changes directly to one button object. How this button fits into the application's interface is visible at all times. Clicking the button allows its functionality to be tested directly. This way, the Lively Kernel makes fast feedback possible, especially during the development of graphical applications.

Programmers' changes to objects can turn out to be inappropriate. Programmers can, for example, accidentally change positions or connect the wrong objects when manipulating applications with mouse interactions. They might try a couple of different alternatives such as different colors and layouts, only to realize that an earlier state was more appealing. Similarly, programmers might learn in hindsight that making a change to an object's scripts introduced an error or impacts the application's performance. They might make a mistake in a code snippet, which then results in the manipulation of many objects. At the same time, it should be noted that problematic changes can also be introduced when code is evaluated for the purpose of testing behavior. In this case the intent is not to make permanent changes to the (current) state (of the program).

However, when changes turn out to be problematic, programmers often need to undo them manually. The Lively Kernel does not provide an undo for changes to objects. This is especially at odds with the Lively Kernel's support for trying ideas right away. In other words, developers are able to make changes directly and receive immediate feedback, but do not get support when such changes turn out to be inappropriate. Thus, to recover a previous development state, programmers often need to manually reset the state to how it previously was—probably using the same tools the changes were initially made with. Furthermore, this potentially involves multiple properties of multiple objects changed by multiple developer actions.

The Lively Kernel provides tools to commit and load versions of objects. In case such commits exist, programmers can load earlier versions of objects to re-establish previous states. Nevertheless, depending on how far the latest version is from the actually desired state, manual changes might still be necessary. To keep the effort to reestablish any previous state low, programmers would need to commit many versions. However, this contradicts the goal because committing many versions is also a significant effort. Some commits would be made only to protect intermediate states and not to share and document results. Especially when the preserved versions should be usable and documented, programmers would be required to test and describe many versions.

In summary, recovering previous states of objects in the Lively Kernel is currently a significant effort for programmers. They either have to manually reset changed state or need to take time-consuming precautionary actions.

A typical approach to implementing multi-level undo for the changes to the application state is the Command pattern (Gamma et al. 1995). The Command pattern packages changes into actions. These actions can then be recorded to be able to subsequently undo them. This requires developers to implement undo operations for all possible actions. Therefore, an implementation of the Command pattern—even when limited to the Lively Kernel tools that manipulate objects—would be rather comprehensive. Furthermore, using the Command pattern requires developers to follow the pattern when implementing new tools. The Command pattern is entirely impractical for undoing the effects of evaluating arbitrary code from the Lively Kernel's workspaces and editors.

Worlds (Warth et al. 2011; Warth 2009), in contrast, is a more generic approach for controlling the scope of side effects. Code is executed in world objects, which capture all side effects. Worlds can then be used to run code with particular sets of changes. Developers could create new worlds for all their actions and the discard worlds to return to previous states when necessary. Therefore, it still requires programmers to explicitly take precautionary actions, similar to version control systems. In addition, the implementation of Worlds in JavaScript is not yet practical. For example, it currently prevents garbage collection.

CoExist (Steinert et al. 2012; Steinert and Hirschfeld 2014) provides automatic recovery support without requiring developers to take precautionary actions. CoExist automatically records versions for every change and, thereby, provides a fine-grained history of intermediate development states. Programmers can review the changes chronologically, examine the impact each change had, and reestablish previous versions. However, CoExist currently recognizes only changes made to the source code of classes. Its versions do not include the state of objects.

We propose an approach for versioning the entire state of programming systems as a basis for automatic recovery support. In particular, we introduce an approach to preserving and managing versions of all objects using alternative, *version-aware* references. Version-aware references are alternative references as they refer to multiple versions of objects. They resolve transparently to stand in for particular versions. Versions of objects are preserved together, so that version-aware references can be resolved transitively to the state of a particular moment. For this to be practical, versions of objects are kept in the application memory and the state of all versions is preserved incrementally on writes. The version-aware references can be changed to a specific version without significantly interrupting program execution. For this implementation, the version-aware references select the current version dynamically instead of being hard-wired to specific versions.

We implemented our approach in JavaScript. The implementation does not require adaptions to established execution engines. Proxies (Cutsem and Miller 2013; Ecma/TS39 2014) are used to implement version-aware references:

conventional references point to the proxies and the proxies delegate all object interactions transparently to particular versions of objects. Source transformations introduce proxies consistently for all objects. Therefore, programmers do not need to adapt their programs manually.

This approach supports fine-grained histories of development states. Not every state can be reestablished—only versions that have been preserved. In this way the presented solution is a basis for recovery support that continuously preserves versions.

2 Background

This chapter describes the approach to prototype-based programming, the Lively Kernel, and CoExist. These works are the background of this thesis as we introduce an approach for providing CoExist-like recovery support in prototype-based programming systems, which we implemented for the Lively Kernel.

2.1 Prototype-Based Programming

Prototype-based programming is object-oriented programming in which applications are created directly with objects, without requiring developers to define classes first. Self, JavaScript, and Kevo (Taivalsaari 1992) are prototype-based programming languages. Many end-user programming systems such as Scratch (Maloney et al. 2010), Etoys (Kay 2005), and Fabrik (Ingalls et al. 1988) also enable users to express programs using objects.

Prototype-based programming makes it possible to build applications from certain objects. This is the fundamental difference to the class-based style of object-oriented programming, in which programs are expressed with classes. Each part of a prototype-based program has a particular state.

There are different advantages associated with this kind of programming:

- Taivalsaari (1996) and Lieberman (1986) suggest that it might be easier for programmers to understand concrete examples than to grasp abstract classes. A concrete example provides particular values for its state and, in the case of objects with a visual appearance, can be actually seen.
- Ungar and Smith (1987) and Borning (1986) describe how prototype-based programming makes it easier to introduce one-of-a-kind objects with their own structure or behavior.
- Borning (1986) and Maloney and Smith (1995) argue that especially in the case of editing, visual objects can be made more concrete with prototypes. Instead of writing code that describes the appearance of objects, programmers can manipulate visual objects directly. Programmers can, for example, use the mouse to

manipulate properties, like the size and position, or to combine multiple elements. This way, programmers always see intermediate states instead of only receiving feedback on explicit test runs in between edit-compile-load cycles.

Editing Graphical Objects at Runtime Many prototype-based programming systems, including the examples given in this section, allow manipulating objects at runtime. Scratch, Etoys, Fabrik, the Lively Kernel, and Self all provide tools dedicated to manipulating graphical objects directly. Such graphical objects range from basic objects like primitive shapes to complete applications like presentation software or programming tools. Prototype-based programming, programming at runtime, and direct manipulation of graphical objects seem to be properties that suit each other.

Similar Objects Without Classes Different prototype-based programming systems provide different approaches for creating similar objects. Self and JavaScript incorporate delegation to allow for prototypical inheritance. Objects can inherit state and behavior directly from other objects: each object has a *prototype* to which it delegates whenever looking up a property if the object itself yields no results. In Self the prototype of an object is set when objects are cloned: The clone's prototype is the object it was cloned from.

In JavaScript objects are created from constructor functions. The constructor function's prototype becomes the prototype of created objects. Kevo, in contrast, does not incorporate this notion of prototypical inheritance. It provides *concatenation* for incremental modification of objects (Taivalsaari 1995). Objects are copied to create objects with the same state and behavior as existing objects. These objects are self-contained. This means, any changes made to an object only affect this object alone. A particular object can only be changed directly and this cannot be carried out by changing any other object. To adapt many objects at once, programmers can use so-called *module operations* in Kevo. Module operations are evaluated for groups of objects.

2.2 The Lively Kernel

The Lively Kernel is a programming system in the tradition of Smalltalk and Self. Development happens at runtime. It incorporates tools and techniques that allow it to be completely self-sufficient. Thus, programmers can create versions of the Lively Kernel with the Lively Kernel. The Lively Kernel is a browser-based system. It is implemented in JavaScript and renders to Hyper Text Markup Language (HTML).

2.2.1 Programming with Prototypes and Classes

As the Lively Kernel is based in JavaScript, the system and applications are expressed in a prototype-based, object-oriented language that provides prototypical inheritance. At the same time, the Lively Kernel also provides a class system and considerable parts of the system are also expressed using classes.

The Lively Kernel implements Morphic (Maloney and Smith 1995), a framework for developing graphical applications. The graphical objects of this framework are called *mMorphs*. Each morph has a class but can also have object-specific behavior. They can be created by instantiating a class or by copying an existing morph. Morphs are often edited directly and not through adapting existing classes or creating new ones. This way, the Lively Kernel mixes the class-based style with the prototype-based style of object-oriented programming.

The Lively Kernel's copy operation does not establish a prototypical inheritance relationship between the copy and the original. Instead, it creates a full copy of the original morph's properties, including its class. Therefore, even though JavaScript incorporates prototypical inheritance, the Lively Kernel encourages programmers to use classes to share behavior among objects.

2.2.2 Direct Manipulation of Morphs

Programmers can change the position of morphs by *dragging*. The composition of morphs can be altered by an alternative dragging called *grabbing*. When a morph is grabbed it can be added to another morph and becomes that morph's submorph. This way, a morph does not have to be a basic shape or simple widget, but can be the interface of any application.

The Lively Kernel provides a set of manipulation tools, called *halos*, as shown in Fig. 1. Developers can bring up these tools for each morph. The different buttons of a morph's halo allow, for example, resizing, rotating, and copying morphs. Other halo buttons open specific tools, which are shown in Fig. 2:

1. The *Inspector* (1) presents all the values that make up a morph's current state. It also has a small code pane at the bottom that can be used to manipulate the morph's properties programmatically.
2. The *Style Editor* (2) allows manipulating certain aspects of a morph's visual appearance. Programmers can use it to change, for example, a morph's color, border width, or the layout of its submorphs.
3. The *Object Editor* (3) is a tool to edit the object-specific behavior of morphs. It shows all scripts of a particular morph and allows programmers to add, remove, and edit scripts.

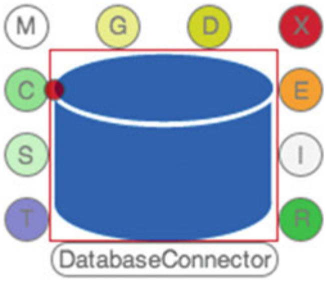


Fig. 1 The halo buttons of a basic morph

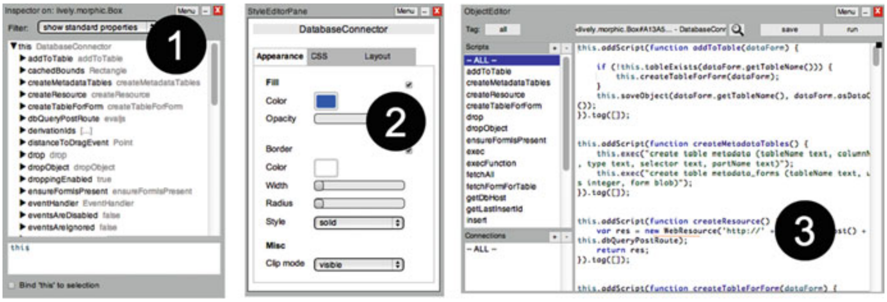


Fig. 2 Three tools of Lively Kernel to manipulate morphs: the inspector, the style editor, and the object editor

2.2.3 Saving Morphs to the Shared Parts Bin Repository

A related tool is the Lively Kernel's *Parts Bin* (Lincke et al. 2012), an object repository to commit and load specific versions of morphs. Morphs saved to the Parts Bin are called *parts* to emphasize the ability to reuse any of the morphs in the Parts Bin for other morphic applications. Figure 3 shows the Parts Bin, opened on the *Tools* category, which includes both the Style Editor and the Object Editor. Both these tools are examples of graphical applications developed from available parts. Their functionality is expressed in scripts and they are available to users through the Parts Bin.

The root of the scene graph of visible morphs is called a *world*. Worlds are not shared via the Parts Bin, but can be saved as Web pages. A world stores the state of all visible morphs when saved and that state can be reloaded with the world.

2.3 *CoExist*

CoExist preserves fast and easy access to previous development states (Krahn et al. 2009). It is based on the insight that the risk for tedious recovery is caused by the loss of immediate access to previous development states. With every change,

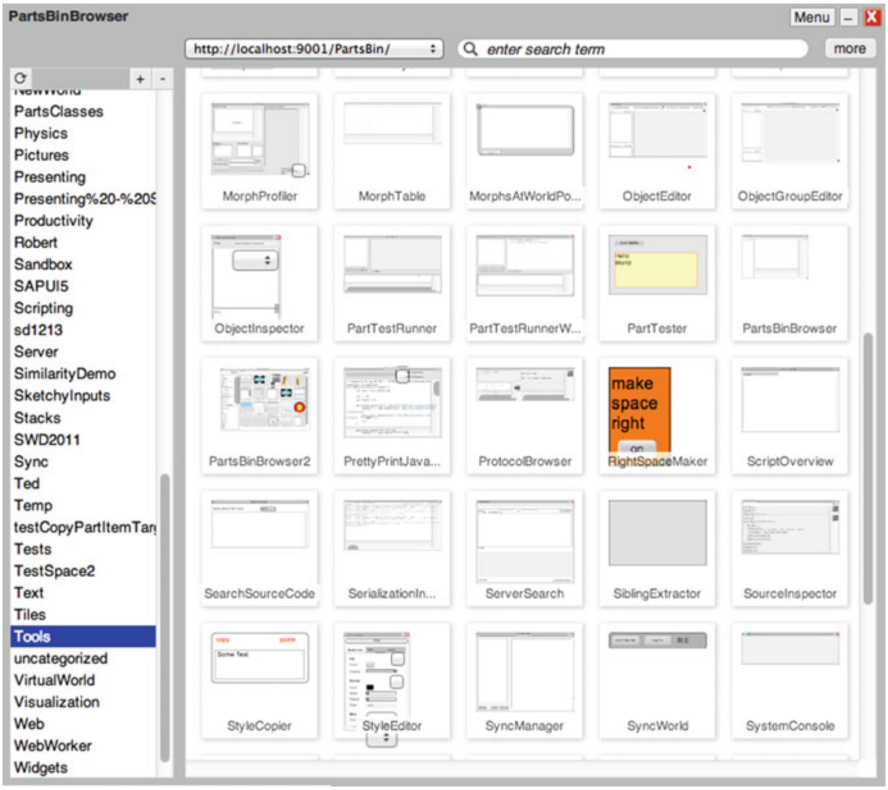


Fig. 3 The Lively Kernel's parts bin opened on the *Tools* category

the previous version is lost, unless it has been saved explicitly. This version, however, can be of value in future development states, when, for example, an idea turns out inappropriate. For that reason, CoExist creates a new version for every change to the code base. Users can rapidly switch versions or can access multiple versions next to each other. CoExist thus gives users the impression that development versions *co-exist*. Figure 4 illustrates some of the main user interfaces concepts of CoExist. It contributes the following concepts and tools:

- *Continuous Versioning* creates new versions in the background based on the structure of programs. It enables programmers to go back to a previous development state and to start over, which will implicitly create a new branch of versions.
- *User Interface Concepts* support browsing and exploring version information as well as quickly identifying a version of interest. Two different tools are provided. First, the version bar highlights version items that match the currently selected source code element. Hovering over items will display additional information such as the kind of modification, the affected elements, or the actual change performed. Second, the version browser allows for exploring multiple

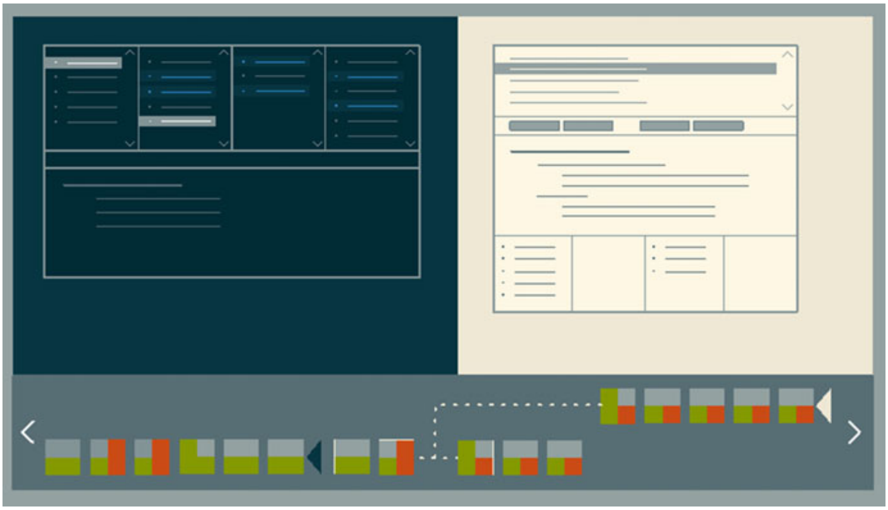


Fig. 4 Conceptual figure of CoExist featuring continuous versioning, running tests and recording the results in the background, and side by side exploring and editing of multiple versions

versions at a glance. The version browser displays basic version information in a table view, which allows for quickly scanning the history.

- *Additional Environments* to explore static and dynamic information of previous development states next to the current set of tools. Opening an addition environment is useful, when, for example, the programmer suddenly becomes curious about how certain parts of the source code looked previously or how certain effects were achieved. The additional environments also allow for running and debugging programs. By way of these, users are capable of efficiently recovering knowledge from previous versions, which avoids the need for a precise understanding of every detail before making any changes.
- *Continuous and Back-In-Time Analysis* for test cases and other computations. CoExist continuously runs analysis programs for newly created versions. As a default, it runs test cases to automatically assess the quality of the change made. The test result for a version is recorded and presented in the corresponding item of the version bar (left of Fig. 4). The user can also run other analyses such as performance measurements. In addition to its continuous analysis features, CoExist provides full access to version objects and offers a programming interface to run code in the context of a particular version. So, whenever programmers become interested in the impact of their changes, they can easily analyze them in various respects. This allows programmers to ignore these aspects of programming at other times.
- *Reassembling of Changes* for sharing independent improvements in separate commits. Users can extract selected changes to a new branch, test the result, and commit the achieved increment.

With CoExist, programmers can change source code without worrying about the possibility of making an error. They can rely on tools that will help with whatever their explorations will turn up. They no longer have to follow certain best practices in order to keep recovery costs low.

3 Motivation

In the Lively Kernel, programmers can create applications by manipulating and composing graphical parts. This section presents the development of such parts and related recovery needs by example.

3.1 Part Development by Example

To exemplify how developers work directly on objects in the Lively Kernel, we will outline how a Lively Kernel user adds a new feature to the Object Editor.

The editor has been developed by composing and editing graphical objects. Thus, the user does not adapt any source code modules to change the editor, but rather manipulates objects directly.

In this example, the user adds a magnifier tool to the Object Editor. The magnifier tool helps find the editor's target, which is the object the editor currently presents scripts for. Implementing the new feature requires creating a new button morph and adding it to the editor, as shown in Fig. 5. The magnifier button has two features:

1. When a programmer hovers over the button, the Object Editor's current target is highlighted with a rectangular overlay.
2. When a programmer clicks the button, the current target selection is revoked and the programmer can select the new target of the editor.

The following description covers the first of the two features, which is also shown in Fig. 6 for an Object Editor currently targeting the character of a game.

Manipulating the Button Morph Before implementing the button's behavior, the user first creates the button and manipulates its visual appearance. Figure 7 shows the steps in which the button is manipulated. A basic button, as visible in (1), can be found in the Parts Bin repository. In (2), the user resizes the button and gives it a square extent using the *Resize* halo button. Next, the user loads an image showing a magnifier icon. Using drag and drop, the image is added to the button in (3). Dropping one morph onto another connects the two morphs. Moving the button around will then move the image accordingly. Finally, the user adds the result of these manipulations, visible in (4), to the Object Editor.

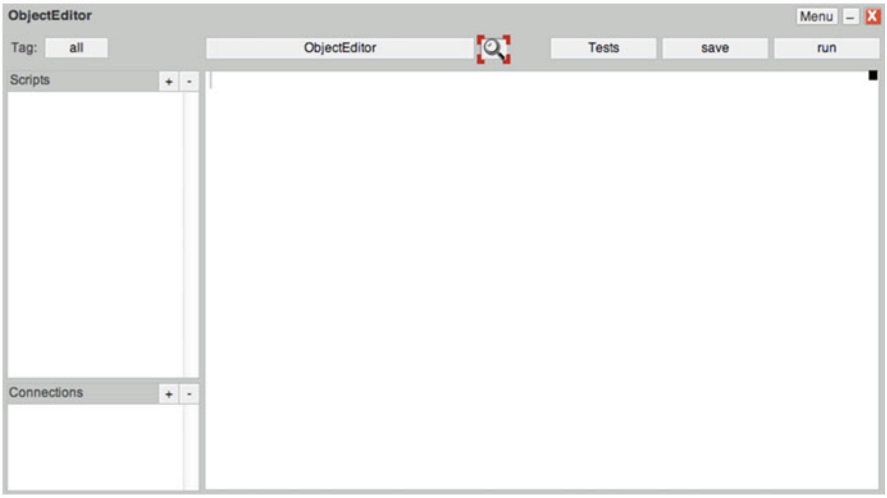


Fig. 5 The object editor's magnifier button highlighted with a red outline

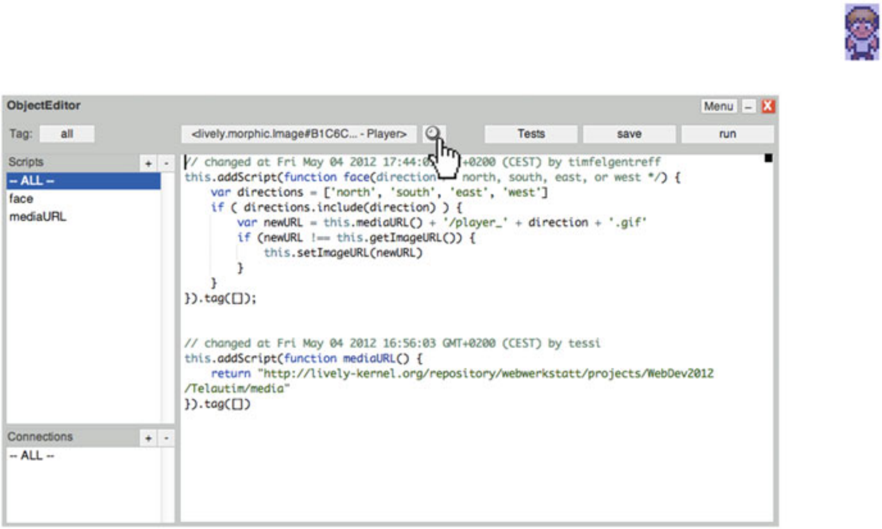


Fig. 6 The object editor's magnifier button as it highlights the editor's target

All these changes to the state of the objects are made directly to the objects themselves: the button morph, the magnifier image morph, and the editor morph. When programmers edit parts in this way, they often see the effects of their actions immediately. For example, when adding the new button to the Object Editor, the button is visible at all times. Programmers do not need to run any code to see and test the button.

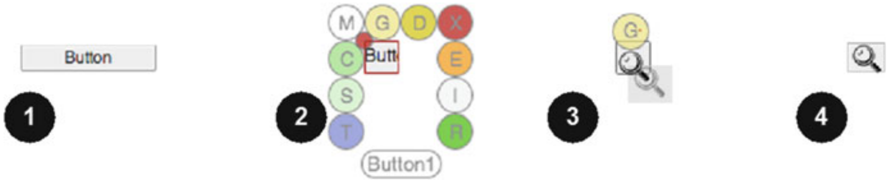


Fig. 7 Directly manipulating a button morph

Scripting the Button Morph Now it is to the user to manipulate the button's behavior. The user adds scripts to the button that lay a translucent rectangle over the current target. In particular, the button receives two scripts: `onMouseMove` and `onMouseOut`. The implementation of the behavior includes the following:

- The button holds a semitransparent rectangle morph.
- When the mouse enters the button (`onMouseMove`), the button resizes and adds the rectangle to the Lively Kernel world at the position of the target.
- When the mouse leaves the button (`onMouseOut`), the button removes the rectangle from the world again.

The Lively Kernel's scripting tools allow evaluating code in the context of their target objects. Hence, when programmers want to test a script or even just specific lines of code, they can try the behavior directly for the actual target.

3.2 Recovery Needs When Developing Parts

While manipulating objects directly, developers might make changes that they later want to undo. For example:

- **Accidental changes to state**: The user could accidentally move a morph and, thereby, change a carefully arranged layout. Similarly, a meaningful state can be lost when a morph, for example the new button, is accidentally removed from the world.
- **Inappropriate changes through direct manipulation**: The user could make changes to the size, position, and colors of morphs to fine-tune the visual appearance of the editor's interface, only to decide later that a particular intermediate version was more appealing.
- **Accidental changes to scripts**: The user could introduce a typographical error or accidentally remove a script. Moreover, editing a script could introduce a defect or a decrease in performance.
- **Inappropriate changes through scripts**: The user could make a mistake in a workspace snippet that is intended to manipulate morph properties programmatically. Such a snippet can change many properties of many objects.

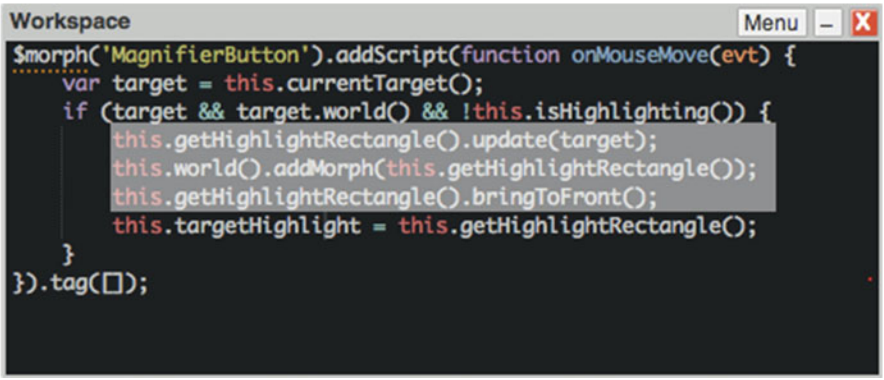


Fig. 8 The button's `onMouseMove` script with a text selection

Explorative Script Evaluation Undesirable changes can also be introduced when a programmer explores the behavior of objects by evaluating scripts. The Object Editor allows evaluating code directly for its target object. While such an evaluation might help to understand the effects of particular code, it might also change the state of the objects. For example, the user could be working on the button's `onMouseMove` script and could evaluate a few lines of code to quickly test them. These lines, as shown in Fig. 8, would add the rectangle to the editor's current target. But evaluating the selected lines would, however, neither check the conditions usually checked above nor set the state usually set below the selected lines. Therefore, evaluating this selection makes it possible to test the highlighting behavior, and, at the same time, leaves the system in a state it normally would not be in.

The examples show that there are many situations in which the user might want to undo previous actions. In programming systems like the Lively Kernel, where programmers work on objects, changes are always made to the state of objects. Functions are properties of objects. Even classes and modules are objects.

For example, evaluating the text selection in Fig. 8 changes the world object's state. The world object now has one more submorph, as shown in Fig. 9. Thus, the world's collection of submorphs is changed.

4 Object Versioning

This section introduces our approach to preserving access to previous states in systems like the Lively Kernel. The approach is based on alternative, version-aware references that manage versions of objects transparently. This section also presents a design that allows implementing version-aware references using proxies.

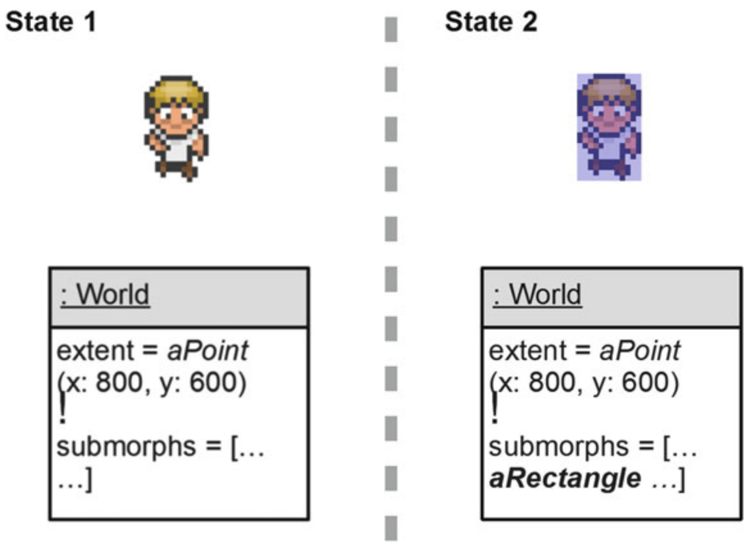


Fig. 9 Adding a submorph changes the state of a morph

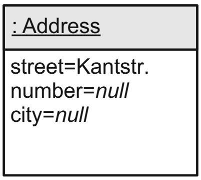


Fig. 10 An address object with three properties

4.1 Version-Aware References

In different versions of a system, objects have different states.

4.1.1 Versions of Objects

An object could represent an address. The state of such an *address* object could be as shown in Fig. 10.

If values are assigned to the city and number fields of the address object, the object's state is changed. As the address object's state is part of the system state, changing the object's state changes the system state as well. If we call the initial state version *v1* and the state after making changes to the object version *v2*, the state of the address object is different in the two versions of the system, as shown in Fig. 11.

To be able to recover previous versions after making changes, the previous states of objects need to be accessible. For this reason, versions of objects are preserved and changes are made to new versions of the objects. A version of an object is, in

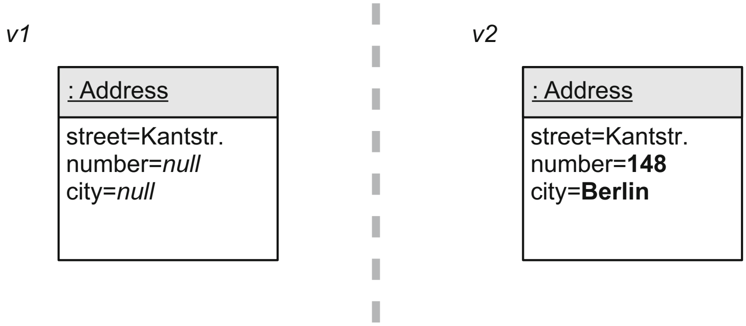


Fig. 11 Two versions of an address object in two versions of the system

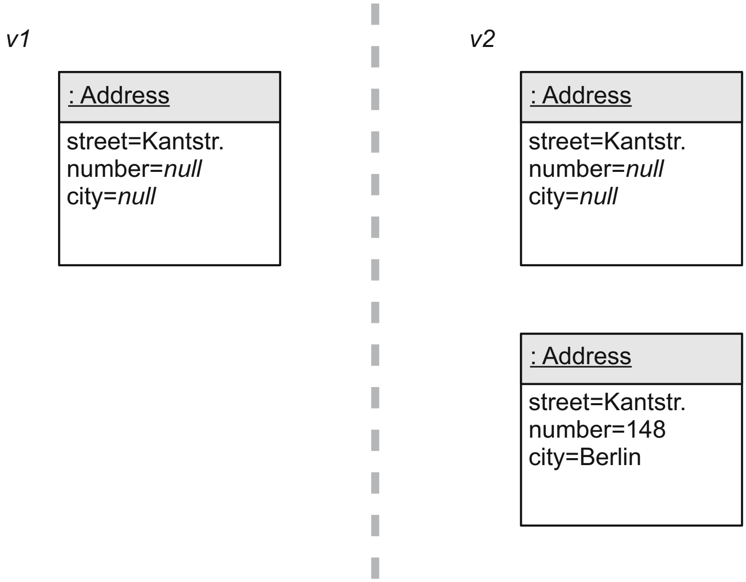


Fig. 12 Preserving the previous version of the address object

the simplest case, a copy of an object. When the address object is changed in version *v2* of the system, the system does not change the original address object but rather the copy.

As shown in Fig. 12, there are now two versions of the address objects in version *v2* of the system. One of the objects holds the original state, while the other holds the state the object should have in version *v2* of the system. The two objects contain no information indicating to which version of the system they belong. They also do not store any information showing that one object is a copy of the other.

At the same time, references to objects remain unchanged. For example, there could have been a person object referring to the address object. This reference

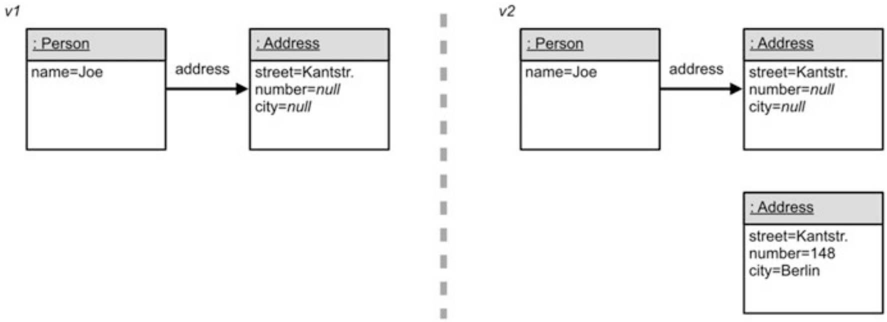


Fig. 13 A reference refers to the previous version of the address object

would still be referring to the original address object, even in version *v2* of the system, as shown in Fig. 13. Even after adding values to the fields of the address object, the following statement would still hold true when `aPerson` refers to the person object:

```
aPerson.address.city === null
```

4.1.2 Version-Aware References

Our approach uses *version-aware references*. Version-aware references recognize the available versions of an object and always resolve to one of these. Furthermore, version-aware references know which object version belongs to which system version. None of the versions is hard-wired to be the active version. Instead, the version-aware references resolve dynamically to the correct version using context information.

Apart from that, the version-aware references behave like ordinary references. They can be assigned to variables and object fields and are thus shared.

When the person object uses a version-aware reference to refer to its address property, it can resolve to the versions of its address object. The version-aware reference knows both versions of the address object. In version *v2* of the system, it resolves to the second version of the object, as shown in Fig. 14.

In the same way, multiple version-aware references can be resolved as one path through a graph of versions. The version-aware references all choose versions of objects that belong to the same system state and, thereby, form the object graph of that state.

Figure 15 shows an object graph that incorporates the previous example. The previously presented person object is a company object's CEO property. While the example shows that version *v2* is active, it also indicates a version *v1* and a version *v2* of the system. In version *v1*, the company's CEO has incomplete address information. In version *v3*, the company has a different CEO.

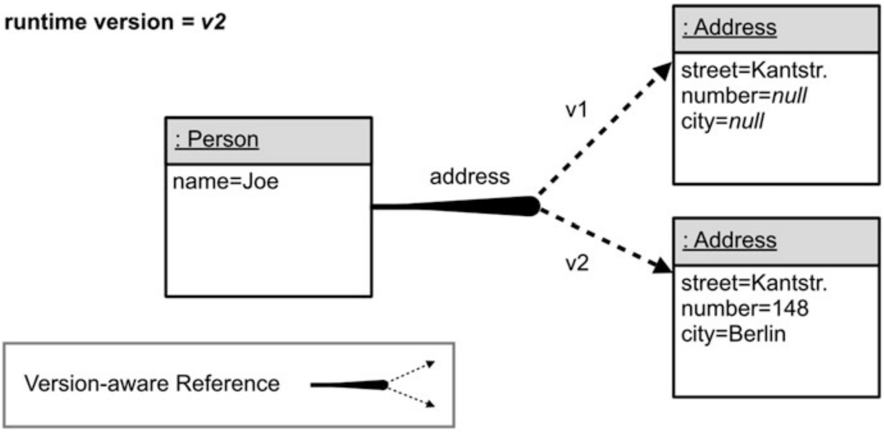


Fig. 14 A version-aware reference relates a person object to two versions of its address property

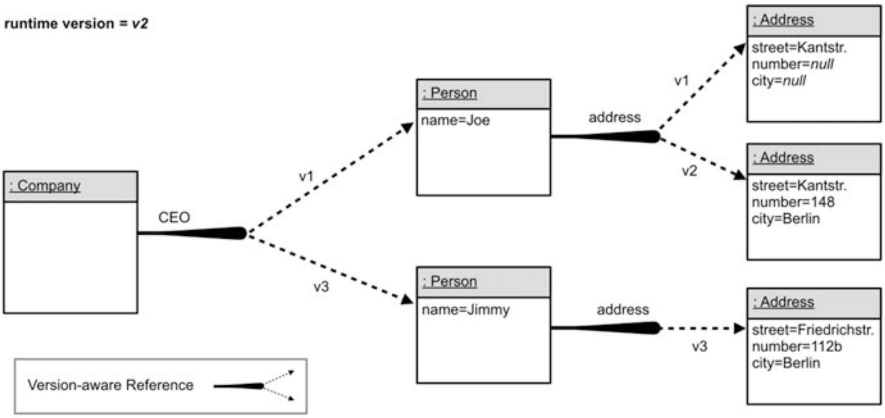


Fig. 15 An object graph with version-aware references

4.1.3 Versions of the System

To establish different versions of the system, the version-aware references have to resolve to different versions of objects. The version-aware references choose versions dynamically following a *version identifier*. It is only necessary to change this version identifier in order to have version-aware references resolve to other versions of objects. For example, to undo the changes made with version *v2* of the system, the version identifier would need to be set to *v1* again.

Given the example situation from Fig. 15 and given that `aCompany` refers to the company object, the following statement would refer to three different values dependent on the version identifier:

```
aCompany.CEO.address.number
```

Evaluating the statement in version *v1* would return the value null, in version *v2* the value 148, and in version *v3* the value 112b.

The information that one version is the predecessor of another version can be used to resolve to an earlier object version when no current version is available. This allows for creating new versions of objects only when necessary.

The version identifier needs to be accessible to the version-aware references. While it could be available globally, to have a single active version of the system, it could also be scoped more locally such as thread-local or in the dynamic scope of a code block. It should, however, not be changed while multiple version-aware references of an object graph are resolved transitively. Consequently, the version-aware references involved in evaluating the previous example statement should be resolved together for the same version identifier.

To be able to actually reestablish a particular version of the system with our approach, two requirements need to be fulfilled. First, all mutable objects of the programming runtime need to be accessed via version-aware references. Second, the particular version of the system needs to be available. Our approach does not allow reestablishing every state but specific states that have to be preserved. Programmers could preserve versions explicitly or the programming system could do this implicitly. When the programming system automatically preserves versions, each programmer action could implicitly yield a new version of the system. This way, programmers could undo and redo the changes of their actions regardless of whether or not they preserved a version in anticipation of recovery needs.

5 Discussion

The presented approach is incremental, not a stop-the-world approach. The version-aware references make it possible to preserve and reestablish versions of the system without completely halting the program execution. First, the version-aware references resolve dynamically to particular versions based on context information. Only this context information has to be changed to have all references resolve to another version. The version-aware references do not have to be re-configured individually.

Second, versions of the system are preserved incrementally. Instead of saving the state of all objects the moment a version is preserved, new versions of objects are created only when objects change. Before such writes, previous object versions continue to reflect the current state and can be read until they are changed.

5.1 *Using Proxies as Version-Aware References*

We used proxies to implement version-aware references in JavaScript. Instead of actually requiring *alternative references*, proxies are referred to by *ordinary references* and transparently delegate to versions. This way, proxies allow a language-level implementation of version-aware references that works with existing JavaScript engines.

5.1.1 Proxies as Version-Aware References

Figure 16 exemplifies how a proxy implements a version-aware reference in our solution. The proxy connects a person object to the two versions of its address property. The person holds an ordinary reference to the proxy in its address slot. The proxy in turn knows which versions are available for the address object.

When the `address` property of the `person` object is accessed, the proxy forwards the access transparently to a version. For example, in version *v2* of the system as indicated in Fig. 16, even if the `address` property is a proxy, reading the proxy's city property returns the string Berlin. Given `aPerson` refers to the `person` object, evaluating the following statement returns true in version *v2*:

```
aPerson.address.city === Berlin
```

The statement does not include any version information. In particular, it does not read a specific version from a table of available versions. Instead, the proxies intercept all object interactions and forward to specific versions transparently. They fulfill three responsibilities:

1. They know which versions are available for a particular object.
2. They choose a particular version among all available dynamically using context information.
3. They forward all interactions transparently to a chosen version.

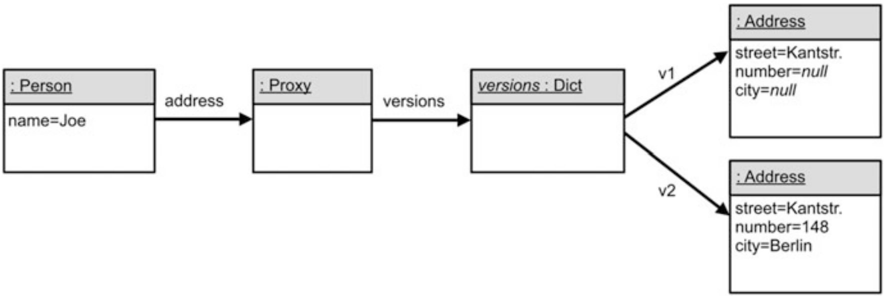


Fig. 16 Using a proxy as version-aware reference to connect a person object to two versions of an address object

The proxies in this design are virtual objects (Van Cutsem and Miller 2010). They do not stand in for specific objects, but can forward intercepted interactions to any object.

5.1.2 Using Proxies Consistently

The proxies need to be used consistently for all mutable states. Ordinary references that usually refer to an object need to refer to the proxy that stands in for the object.

To use proxies consistently, we create and return proxies for all new objects. All expressions that create new objects return proxies for those objects instead. This is achieved by transforming code before it is executed. The source transformations wrap object literals and constructor functions into proxies. The proxies also always return proxies as return values. Thus, when proxied constructors are used, the constructors return proxies for the new objects.

The proxy encapsulates the reference to the initial version of an object. The reference to the proxy is passed around instead. For this reason, all references that would usually point to the same object point to the same proxy. This way, proxies provide object identity. Checks that would usually compare one object to another object now compare one proxy to another proxy.

Only the proxies contain references to the versions of objects. This means that when the proxies are no longer reachable the versions get garbage collected along with the proxies. For example, in the code of Listing 4.1 there temporarily exists a version-aware reference—a proxy—connecting the person object to an address object. However, before a version of the system is preserved the reference gets deleted. The address object is not required to reestablish either version *1* or version *2* of the system and nothing prevents the garbage collector from reclaiming the proxy for the address object with the address object.

```
var person = {name: "Joe"};

 \\ [preserve first version]
person.address = {street: "Kantstr.",
                  number: "148",
                  city: "Berlin"};

delete person.address;

\\ [preserve second version]
```

Listing 4.1 A newly created object is not preserved with any version

5.1.3 Versions of the System

Proxies delegate to and create versions of an object using a *version of the system*. A version of the system is an object that has a *version identifier*, a *predecessor* version, and a *successor* version. Figure 17 shows three system versions. In the example, version *v2* is the current version of the system.

The current version of the system is accessible to the proxies. Proxies use it to decide to which version of an object they should currently forward. Figure 18 shows a proxy, versions of an object, and versions of the system. In this example, there are two object versions that correspond to the two system versions. The current version of the system is *v2* and, therefore, version *v*2 of the object is the target that the proxy currently forwards to.

As long as the system version stays the same, the proxies forward to the same version of the object. Therefore, an object version is changed only as long as it matches the current system version. *To reestablish the previous version*, the system

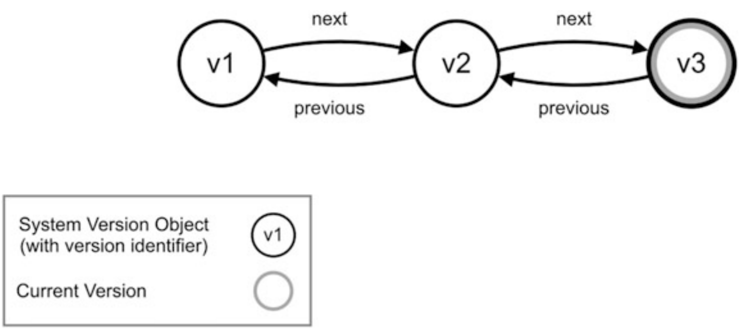


Fig. 17 Four versions of the system

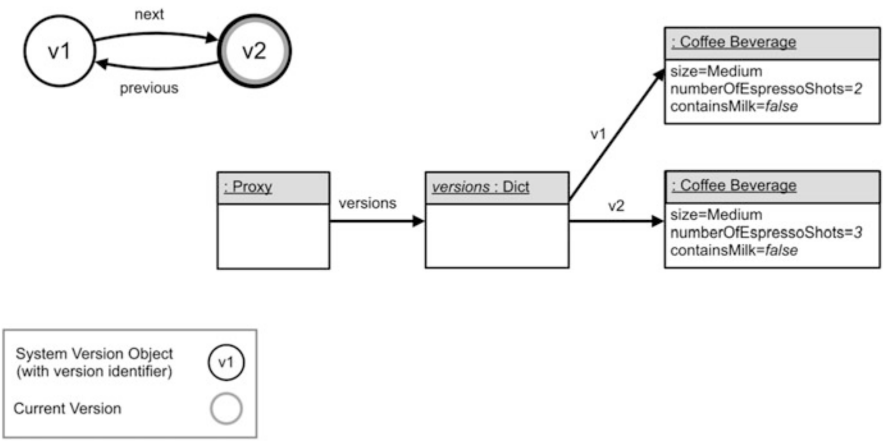


Fig. 18 A proxy with two object versions in context of the system versions

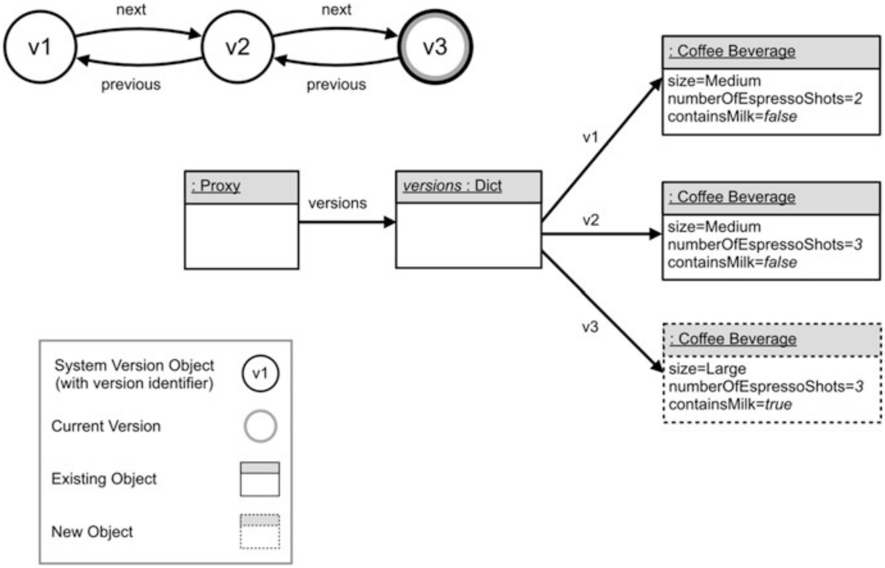


Fig. 19 A new version of an object is created for a new version of the system

version has to be set to its predecessor. In that case, proxies forward interactions to previous versions of the objects. *To preserve the current version*, the system version has to be set to a different version. The proxies forward interactions to other object versions or, when no such version of the object exists, they create new versions.

A situation in which a new version of an object is created is shown in Fig. 19. In a new version *v3* of the system, the proxy intercepts a manipulation but has no object version it can forward to. It therefore copies the most recent version of the object and forwards manipulation request to this new copy.

New versions are only necessary when a proxy is about to delegate manipulations. As long as the state of an object is only read, the proxy reports values from a previous version. The old version of the object still reflects the current state. To create a new version, a proxy copies the most recent previous version of the object.

5.1.4 Limitations

The current design allows for to preserving and reestablishing versions of the system. However, without further components these versions only exist in memory and are not stored to disk. Our current design does not support multiple predecessors or successors. Another limitation of the current design is that the state of previous versions can be changed. New versions of objects are not affected by changes to previous versions, but changes to object versions that have not been copied are present in subsequent versions of the system.

In the future, the versioning might allow for branches and merging. This means that changes to previous states could then be handled in branches that programmers may or may not merge into future versions.

6 Related Work

This section presents two categories of related work:

1. Approaches related to our motivation and, thus, to providing access to previous versions of the system state.
2. Approaches related to our technical solution and, thus, to combining changes into first-class objects that can be used to scope changes.

6.1 Recovering Previous System States

The approaches presented in this section support programmers in recovering previous states without requiring programmers to create snapshots in advance.

6.1.1 CoExist

CoExist (Steinert et al. 2012) provides recovery support through continuous versioning in Squeak/Smalltalk. For each change made to source code, CoExist creates a new version of the system sources, resulting in a fine-grained history of changes. CoExist presents this history in a timeline tool and a dedicated browser. For each version, these tools show the changes, test results, and a screenshot. Developers can recover previous development states, even without taking precautionary actions beforehand. This way, developers can concentrate on implementing their ideas and let CoExist record the required versions that are able to be recovered when necessary.

Both CoExist and our approach of object versioning allow multiple versions of the development state to coexist. With both approaches, preserved versions are part of the program runtime and can be reestablished easily. Currently, CoExist only records versions on the granularity of changes made by developers. CoExist provides much more tool support to find and recover changes from previous versions. However, CoExist recognizes only changes to the source code of classes, while our system preserves the state and behavior of objects.

6.1.2 Back-in-Time Debugging

Back-in-time debuggers (Lewis 2003), also known as *omniscient debuggers*, allow developers to inspect previous program states and step backwards in the control flow to undo the side effects of statements. Approaches for this are based either on logging or replay. In other words, either the debugger records information to be able to recreate particular previous situations, which mainly requires space for the different states, or the debugger re-executes the program up to a particular previous situation, which mainly requires time to re-run the program. While many logging-based approaches introduce significant execution overheads, replay-based approaches have to ensure that the program is re-executed deterministically. This re-execution can be a difficult problem when, for example, programs rely on a state outside of the program runtime, such as the content of files or the state of other programs.

Our approach is more related to logging-based back-in-time debugging. It allows reestablishing a previous state through preserving information. However, back-in-time debuggers need to be able to undo the effects of each statement separately, while our system's versioning granularity is arbitrary and can, for example, correspond to programmer interactions with the system. In general, back-in-time debuggers support a particular development task—debugging—and, thus, are also often only active when a program is started in a separate debugging mode. In contrast, the purpose of object versioning is more comprehensive. We expect object versioning to be active at least during all development tasks and to be enabled at all times if possible.

6.1.3 Software Transactional Memory

Software Transactional Memory (STM) (Shavit and Touitou 1995) captures changes to values in transactions, analogous to database transactions. Each transaction has its own view of the memory, which is unaffected by other concurrently running transactions. Multiple versions of the system state can coexist. The version that is read and written depends on the transaction. Transactions contain a number of program statements that are executed atomically. The changes of a transaction are only permanent when no conflicts occur with other transactions. When conflicts are involved all changes from the transaction are rolled back and undone.

STM and our approach are similar in that multiple versions of the system state can coexist and that a previous state can be reestablished if necessary. However, STM provides concurrency control and an alternative to lock-based synchronization, while our approach provides recovery support to developers when changes turn out to be inappropriate. STM transactions are automatically rolled back when the changes it instigates conflict with changes from other concurrently running transactions. In comparison, our versions are offered to programmers to undo changes when necessary. Programmers can actively decide to undo changes when these, for example, negatively impact the functionality, design, or performance of programs. Our versions of the runtime are also first-class objects, which can be stored in variables and be re-established at any time, while transactions are always created implicitly through particular control structures and committed immediately upon success.

6.2 Dynamically Scoping First-class Groups of Changes

The approaches presented in this section allow combining changes into first-class objects and running code with particular sets of changes.

6.2.1 Worlds

Worlds provide a language construct for controlling the scope of side effects: changes to the state of objects are by default only effective in the *world* in which the changes occurred. Worlds are first-class values and can be used to execute statements with particular side effects being active. A new world can be spawned from an existing world, which establishes a child-parent relationship between the two worlds. Developers can commit changes from a child world to its parent world, thereby extending the scope of the captured side effects. The world approach includes conditions that prevent commits that would potentially introduce inconsistencies.

In comparison, Worlds provides a language construct for experimenting with different states of the system, while object versioning allows for preserving versions of the system to recover previous states. Our approach does not include extensions to the host programming languages and no conditions for combining versions with their predecessor versions, but instead provides a basis for CoExist-like continuous versioning and recovery tools.

Other differences between Worlds and our approach regard implementations. Our implementation in JavaScript does not prevent garbage collection as Worlds does. Further, both use different libraries for source transformations. Our source transformations are faster and do not use JavaScript exceptions.

6.2.2 Object Graph Versioning

Object Graph Versioning (Pluquet et al. 2009) allows programmers to preserve access to previous states of objects. Fields of objects can be marked as *selected* fields. When a snapshot is created, the values of these selected fields are preserved. Therefore, not every state can be reestablished, but only those states that are part of global snapshots. The approach, thus, provides fine-grained control to programmers regarding which fields of which objects should be preserved and when.

The technical solution is similar to our design. Analogous to our proxy-based version-aware references, selected fields do not refer directly to their actual values, but to chained arrays that manage multiple versions of the state of a field and delegate access to the current version transparently. The chained arrays decide which version to retrieve and when to create new versions using a global version identifier. In contrast to our solution, individual fields are versioned and only when programmers explicitly mark them as selected.

Object Graph Versioning aims to support implementing application-specific undo/redo or tools like back-in-time debuggers. In contrast, our approach to object versioning aims to support recovery of previous system states during the development of arbitrary applications.

6.2.3 Context-Oriented Programming

Context-oriented Programming (COP) (Hirschfeld et al. 2008; Appeltauer et al. 2009) adds dedicated language constructs for dynamic behavior variations. Depending on context information, COP allows enabling and disabling layers that contain methods to be executed instead of or around methods of the base programs. Context information can be any information that is computationally accessible. Layers can be enabled and disabled at runtime. Different implementations of COP provide different mechanisms to scope the activation of layers: for example, layers can be activated explicitly for a particular scope or globally for the entire runtime. In ContextJS (Lincke et al. 2011) it is possible to activate layers for specific objects.

In comparison to our approach COP allows activating combinations of layers, while our system executes code using a single active version. In COP layers are independent, while our versions are predecessors and successors of each other.

COP aims at supporting the separation of heterogeneous crosscutting concerns, while object versioning aims at supporting developers with the recovery of previous states. However, Lincke and Hirschfeld (2012) showed that COP can also be used to experiment with changes to a system: developers can implement experimental changes to behavior in layers, not to modularize context-dependent adaptions, but to be able to scope changes dynamically and recover the original system behavior easily. However, this requires programmers to make experiments explicitly. They need to use layers for their adaptions, enable the layers for test runs, and move code from layers back to the base system when experiments are successes and they want to maintain the original modularization of the system. COP also allows only behavior variations, while our approach recognizes changes to both state and behavior.

6.2.4 ChangeBoxes

ChangeBoxes (Denker et al. 2007) is an approach to capturing and scoping changes to a system using first-class entities, called ChangeBoxes. A ChangeBox can contain changes to multiple elements of a software system such as adding a field, removing a method, or renaming a class. The approach does not constrain how changes get grouped into ChangeBoxes, but every change has to be encapsulated by a ChangeBox. Each ChangeBox can be used for setting the set of active changes for the scope of a running process. This way, multiple running processes can view the system differently by using different ChangeBoxes. ChangeBoxes can have ancestor relations and merge changes from multiple ancestors.

With the ancestor relations, ChangeBoxes can be used to review the evolution of systems and to undo changes.

The ChangeBoxes approach is similar to our approach as changes to the system are grouped into first-class objects and these can be used to run code in different versions of the runtime. Furthermore, with both solutions there is no definite notion on how changes are grouped into versions. Our object versioning approach is intended to be used to group changes associated with developer actions and a simple global undo/redo mechanism to undo inappropriate actions is built into our solution. To actually undo changes Change-Boxes, in contrast, is rather tedious (Steinert et al. 2012). Moreover, ChangeBoxes recognizes only changes to the static elements of a software system such as packages, the structure of classes, and methods. Object versioning, in contrast, preserves the state and behavior of objects.

6.2.5 Practical Object-Oriented Back-in-Time Debugging

Practical Object-Oriented Back-in-Time Debugging (Lienhard et al. 2008; Lienhard 2008) is a logging-based approach to back-in-time debugging that uses alternative references to preserve the history of objects. These alternative references, called *aliases*, are actually objects and part of the application memory. These objects contain information about the history and origin of the values stored in fields. Aliases are not passed around, but instead are created for each read or write of a field and for each value passed as parameter. Each alias refers to an actual value, but also to another alias—its *predecessor*—representing the value previously stored by a field and to the alias that was used to create this new alias from—its *origin*. An alias and its origin both refer to the same value, but provide different information on their creation context, which is a particular method. The origin link can be used to follow the object's "flow" through the program. Each alias also records a timestamp on its creation and with this information the predecessor link can be followed to read a value as it was at a particular moment in time.

In comparison, with aliases it is possible to recreate all previous system state and also to retrace the flow of all values, while our system stores only particular versions. Such versions could, for example, correspond to programmer interactions, so that programmers can undo the effects of particular actions easily. Another difference between object versioning with version-aware references and reverse engineering with aliases is the existence of modes. The alias references are intended to be used in explicit debugging sessions, while our version-aware reference are intended to be used at all times.

7 Summary

This work introduced an approach to preserving access to previous states of programming systems such as the Lively Kernel. The approach is based on version-aware references. These references manage different versions of objects transparently. They resolve to one of multiple versions of objects. Thereby, different preserved states can be re-established.

We presented a design for our approach that uses proxies for version-aware references. Instead of actually using alternative references, ordinary references refer to proxies and proxies forward all interactions transparently to the right versions. The design allows implementing version-aware references without any adaptions to existing execution engines—neither for alternative references nor for the garbage collection of versions.

For each object that is created, a proxy is created and returned instead of the object. Thus, references to proxies are passed around and all access goes through the proxies. Moreover, only the proxies refer to the versions of an object. Consequently, the versions of an object are reclaimed together with their proxy by the ordinary garbage collector. Returning proxies for new objects is achieved using source transformations. The program sources are transformed when loaded and do not have to be adapted manually.

References

Appeltauer M, Hirschfeld R, Haupt M, Lincke J, Perscheid M (2009) A comparison of context-oriented programming languages. In: International workshop on context-oriented programming. COP'09. ACM, July 2009, 6:1–6:6

Borning A (1986) Classes versus prototypes in object-oriented languages. In: Proceedings of 1986 ACM fall joint computer conference. ACM'86. IEEE, November 1986, pp 36–40

Cutsem T, Miller MS (2013) Trustworthy proxies: virtualizing objects with invariants. In: Proceedings of the 27th European conference on object-oriented programming. ECOOP'13. Springer, July 2013, pp 154–178

Denker M, Gîrba T, Lienhard A, Nierstrasz O, Renggli L, Zumkehr P (2007) Encapsulating and exploiting change with changeboxes. In: Proceedings of the 2007 international conference on dynamic languages. ICDL'07. ACM, August 2007, pp 25–49

Ecma/TS39 (2014) ECMAScript language specification (Draft for 6th Edition). Published April 27, 2014 (Draft, Revision 24). http://wiki.ecmascript.org/doku.php?id=harmony:specification_drafts. Accessed 11 May 2014

Gamma E, Helm R, Johnson R, Vlissides J (1995) Design patterns: elements of reusable object-oriented software. Addison-Wesley, Reading, MA

Goldberg A, Robson D (1983) Smalltalk-80: the language and its implementation. Addison-Wesley, Reading, MA

Hirschfeld R, Costanza P, Nierstrasz O (2008) Context-oriented programming. J Object Technol 7 (3):125–151

Ingalls D, Wallace S, Chow Y-Y, Ludolph F, Doyle K (1988) Fabrik: a visual programming environment. In: Conference proceedings on object-oriented programming systems, languages and applications. OOPSLA'88. ACM, January 1988, pp 176–190

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. In: Proceedings of the 12th ACM SIGPLAN conference on object-oriented programming systems, languages and applications. OOPSLA'97. ACM, October 1997, pp 318–326

Ingalls D, Palacz K, Uhler S, Taivalsaari A, Mikkonen T (2008) The Lively Kernel—a self-supporting system on a web page. In: Self-sustaining systems. S3. Springer, May 2008, pp 31–50

Kay A (2005) Squeak Etoys authoring and media. Technical report. Published February 2005. http://www.vpri.org/pdf/rn2005002_authoring.pdf. Accessed 7 Mar 2014

Krahn R, Ingalls D, Hirschfeld R, Lincke J, Palacz K (2009) Lively wiki: a development environment for creating and sharing active web content. In: Proceedings of the 5th International symposium on wikis and open collaboration. WikiSym'09. ACM, October 2009, 9:1–9:10

Lewis B (2003) Debugging backwards in time. In: Proceedings of the fifth international workshop on automated debugging. AADEBUG'03. Springer, September 2003, pp 225–235

Lieberman H (1986) Using prototypical objects to implement shared behavior in object-oriented systems. In: Conference proceedings on object-oriented programming systems, languages and applications. OOPLSA'86. ACM, June 1986, pp 214–223

Lienhard A (2008) Dynamic object flow analysis. Ph.D. thesis. University of Bern, December 2008

Lienhard A, Gîrba T, Nierstrasz O (2008) Practical object-oriented back-in-time debugging. In: Proceedings of the 22nd European conference on object-oriented programming. ECOOP'08. Springer, July 2008, pp 592–615

Lincke J, Hirschfeld R (2012) Scoping changes in self-supporting development environments using context-oriented programming. In: Proceedings of the international workshop on context-oriented programming. COP'12. ACM, June 2012, 2:1–2:6

Lincke J, Hirschfeld R (2013) User-evolvable tools in the web. In: Proceedings of the 9th international symposium on open collaboration. WikiSym'13. ACM, August 2013, 19:1–19:8

Lincke J, Appeltauer M, Steinert B, Hirschfeld R (2011) An open implementation for context-oriented layer composition in ContextJS. Sci Comput Program 76(12):1194–1209

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 2012 45th Hawaii international conference on system sciences. HICSS'12. IEEE, January 2012, pp 693–701

Maloney JH, Smith RB (1995) Directness and liveness in the morphic user interface construction environment. In: Proceedings of the 8th annual ACM symposium on user interface and software technology. UIST'95. ACM, December 1995, pp 21–28

Maloney J, Resnick M, Rusk N, Silverman B, Eastmond E (2010) The scratch programming language and environment. Trans Comput Educ 10(4):16:1–16:15

Pluquet F, Langerman S, Wuyts R (2009) Executing code in the past: efficient in-memory object graph versioning. In: Proceedings of the 24th ACM SIGPLAN conference on object oriented programming systems languages and applications. OOPSLA'09. ACM, October 2009, pp 391–408

Shavit N, Touitou D (1995) Software transactional memory. In: Proceedings of the fourteenth annual ACM symposium on principles of distributed computing. PODC'95. ACM, June 1995, pp 204–213

Steinert B, Hirschfeld R (2014) How to compare performance in program design activities: towards an empirical evaluation of CoExist. In: Design thinking research. Understanding innovation. Springer, January 2014, pp 219–238

Steinert B, Cassou D, Hirschfeld R. CoExist: overcoming aversion to change. In: Proceedings of the 8th symposium on dynamic languages. DLS'12. ACM, January 2012, pp 107–118

Taivalsaari A (1992) Kevo, a prototype-based object-oriented language based on concatenation and modules operations. Technical report LACIR 92-02. University of Victoria

Taivalsaari A (1995) Delegation versus concatenation or cloning is inheritance too. In: SIGPLAN OOPS messenger 6.3, July 1995, pp 20–49. Issn: 1055-6400

Taivalsaari A (1996) Classes vs. prototypes-some philosophical and historical observations. J Object Orient Prog 10(7):44–50

Ungar D, Smith RB (1987) Self: the power of simplicity. In: Conference proceedings on object-oriented programming systems, languages and applications. OOPSLA'87. ACM, December 1987, pp 227–242

Ungar D, Smith RB (2007) Self. In: Proceedings of the third ACM SIGPLAN conference on history of programming languages. HOPL III. ACM, San Diego, CA, June 2007, 9:1–9:50

Van Cutsem T, Miller MS (2010) Proxies: design principles for robust object-oriented intercession APIs. SIGPLAN Not 45(12):59–72

Warth A (2009) Experimenting with programming language. Ph.D. thesis. University of California, Los Angeles, December 2009

Warth A, Ohshima Y, Kaehler T, Kay A (2011) Worlds: controlling the scope of side effects. In: Proceedings of the 25th European conference on object-oriented programming. ECOOP'11. Springer, July 2011, pp 179–203

Connecting Designing and Engineering Activities III

Thomas Beyhl and Holger Giese

Abstract Nowadays, companies implement innovation processes or outsource them to external consulting companies to gain a competitive business advantage. The methodology of Design Thinking is one example for such an innovation process that enables the creation of innovative products or services, which make sense to people and for people, are likely to become a sustainable business model as well as are functionally possible in the foreseeable future.

But, innovators and engineers are seldom the same people, what makes documenting innovation projects with a subsequent information handover inevitable. In practice, this information handover seldom goes smoothly due to missing, incomplete, or not traceable documentation of innovation projects. The situation becomes even worse when important design rationales, design paths and design alternatives are not retrievable anymore, because the missing information may lead to a realization of the innovation that was not intended by the innovators. The retrieval of design paths, design rationales, and design alternatives requires high manual effort, if at all possible.

In this chapter, we present a recovery approach that eases the retrieval of design artifacts by recovering design paths before the actual retrieval happens. For that purpose, our recovery approach employs recovery modules that implement knowledge extraction procedures and recovery algorithms. We evaluate our recovery approach using inventory documentation collected in educational Design Thinking settings.

1 Introduction

Nowadays, companies implement innovation processes or outsource them to external consulting companies to gain a competitive business advantage. The methodology of Design Thinking (Plattner et al. 2009) is one example for such an

T. Beyhl (✉) • H. Giese
Hasso Plattner Institute for Software Systems Engineering, Campus Griebnitzsee, 14482 Potsdam, Germany
e-mail: thomas.beyhl@hpi.de; holger.giese@hpi.de

H. Plattner et al. (eds.), *Design Thinking Research*, Understanding Innovation,
DOI 10.1007/978-3-319-19641-1_16

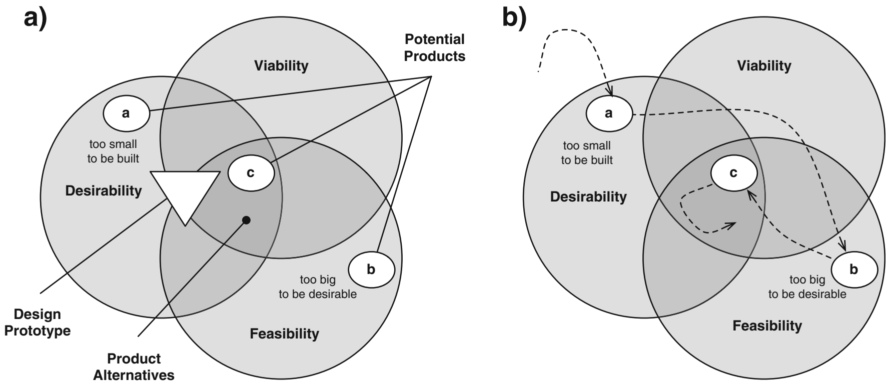


Fig. 1 Three dimensions of Design Thinking, cf. Beyhl et al. (2013c), Brown (2009)

innovation process that enables the creation of innovative products or services. According to Brown (2009) innovative products and services *"make sense to people and for people"* (desirability), *"are likely to become a sustainable business model"* (viability) as well as are *"functionally possible in the foreseeable future"* (feasibility) as depicted by the left hand side of Fig. 1. Therefore, Design Thinking is often considered as incubator for innovative products and services (innovative ideas for short).

But, innovators and engineers are seldom the same people, what makes the transition from designing an innovative idea to engineering it a challenging task. Thus, documenting innovation projects with a subsequent information handover is inevitable. In practice, this information handover seldom goes smoothly due to missing, incomplete, or not traceable documentation of innovation projects. For example, a design prototype often focuses on conveying the overall idea instead of serving as a blueprint for manufacturing. Therefore, design prototypes often cover the three dimensions of desirability, viability, and feasibility only partially. The situation becomes even worse when important design rationales, design paths and design alternatives are not retrievable anymore, because the missing information may lead to a realization of the innovation that was not intended by innovators. We support this retrieval by recovering the design path including design rationales and design alternatives before the actual retrieval happens and, therefore, ease the retrieval task later on.

To support the retrieval of relevant design artifacts, innovation processes need to support traceability (Beyhl et al. 2013c), as depicted by the right hand side of Fig. 1. Traceability is the *"ability to describe and follow the life of a requirement in both a forwards and backwards direction"* (Gotel and Finkelstein 1997). Gotel introduces the concepts of signs, tracks, and traces for traceability (Gotel and Morris 2011). According to Gotel (Gotel and Morris 2011), a sign is *"an identifying mark made by, or associated with for a particular purpose, an animate or inanimate object"*. In Design Thinking, Design Thinkers often create analog artifacts such as post-its, which are captured by digital artifacts such as photographs. Furthermore, analog artifacts are seldom transcribed for further processing, although these artifacts embody important design rationales, end-user feedback, and ideas that need to be retrieved later on, when the innovative idea is made ready for the market. In Design Thinking, these artifacts are signed entities that need to be associated with artificial signs to be traceable, because they do not consist of natural signs. Design Thinkers move these artifacts between different contexts and working states, which are often captured in terms of photographs as well. These movements constitute a track that represents the Design Thinker's journey. Therefore, Design Thinkers are considered as sign and track makers. Engineers primarily use the created signs and tracks. Therefore, engineers are considered as primary trace users. They primarily trace in backward direction from the final documentation to earlier milestones to look up design rationales and design alternatives. Design Thinkers are secondary trace users, who use the traces for synthesis and reflection or when a follow-up Design Thinking project should be kicked off.

However, as in other disciplines (Arkley and Riddle 2005) traceability is often considered as beneficial by trace users only and as obstructive by sign and track makers. We observed that asking Design Thinkers to explicitly leave signs to create a track for the sole purpose of design documentation (Menning et al. 2014) and later retrieval does not work well. Thus, a manual retrieval of relevant design artifacts is quite cumbersome and requires high effort, if possible at all. To ease the retrieval, the design path including design rationales and design alternatives needs to be recovered from daily working documents of Design Thinkers without explicitly asking Design Thinkers to document their progress using a certain documentation scheme.

In this chapter, we present a recovery approach (Sect. 2) that enables the reconstruction of the design path by employing recovery modules (Sect. 3) that implement arbitrary knowledge extraction procedures and recovery algorithms to turn implicit knowledge into explicit knowledge. In contrast to this implicit knowledge, engineers and Design Thinkers can query this explicit knowledge. We evaluate the recovered design path using inventory documentation collected in educational Design Thinking settings (Sect. 4) and delimit our approach from related work (Sect. 5) before concluding our approach (Sect. 6).

2 Recovery Approach

Figure 2 gives an overview about our recovery approach. The main idea of our recovery approach is an active repository (Beyhl et al. 2013a; Beyhl and Giese 2015). An active repository is a repository that is steadily analyzed by recovery modules. In doing so, the degree of retrievable information stored within the repository is increased and information that would not be included explicitly in the repository without analysis can be retrieved easily.

In Design Thinking, the repository consists of Design Thinking project documentation, e.g. photographs of whiteboards and text documents that, e.g., describe

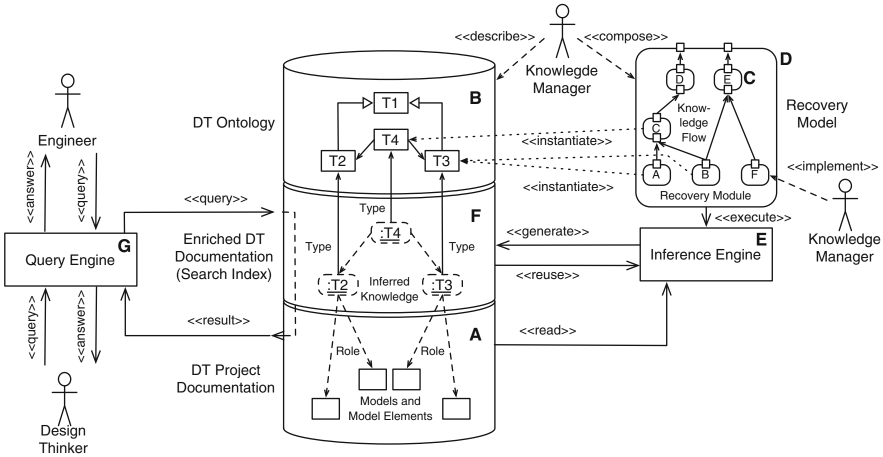


Fig. 2 Overview about recovery approach

point of views or innovative ideas. Furthermore, in Design Thinking the recovery modules reconstruct the design path including design rationales and design alternatives.

Several manual and automatic activities are required to enable the recovery of the design path. First, Design Thinking documentation artifacts (see A in Fig. 2) need to be captured, e.g. with the help of file shares or Design Thinking project documentation tools such as ProjectZoom (Beyhl and Giese 2015). Second, an ontology of Design Thinking artifacts and relevant design knowledge (Design Thinking ontology for short) (see B in Fig. 2) is required that describes, which kind of knowledge have to be recovered and, therefore, can be retrieved later on when recovered. Third, for each element in the Design Thinking ontology at least one recovery module (see C in Fig. 2) has to be implemented, which recovers the knowledge. Further, the knowledge recovered by one recovery module can be exploited by dependent recovery modules to recover additional knowledge based on already recovered knowledge. Thus, all recovery modules together constitute a recovery model (see D in Fig. 2) in terms of a directed acyclic graph. Fourth, the inference engine (see E in Fig. 2) executes the recovery model and, thereby, triggers the execution of the recovery modules. Due to the execution of the recovery modules, the Design Thinking ontology is instantiated in terms of a search index (see F in Fig. 2). Afterwards, the additionally inferred knowledge can be queried with the help of a query engine (see G in Fig. 2).

In this chapter, we build on the results of our previous work how to capture design artifacts (Beyhl et al. 2013a; Beyhl and Giese 2015). Now, we focus on the creation of the Design Thinking ontology and the creation and execution of the recovery model. The Design Thinking ontology describes the kinds of artifacts included in Design Thinking project documentation as well as the kind of knowledge and its relationships that need to be recovered. But, the Design Thinking ontology does not describe how to recover the required knowledge. Therefore, recovery modules are employed that need to be specified separately to recover the knowledge. In other words, the recovery modules instantiate the Design Thinking ontology and, therefore, enable the retrieval of Design Thinking related knowledge.

3 Design Thinking Documentation Recovery

In this section, we present our Design Thinking ontology and introduce recovery modules that instantiate the Design Thinking ontology. In the following sections, we distinguish between different categories of knowledge: meta-data, stakeholder related knowledge, methodology (process step, activity, technique) related knowledge, and design rationale related knowledge.

3.1 Design Thinking Ontology

Ontologies define a common terminology of domain-specific terms and their relationships. Gruber defines ontologies as "*explicit specification of a conceptualization*" (Gruber 1994). Further, Gruber states that *"in such an ontology, definitions associate the names of entities in the universe of discourse (e.g., classes, relations, functions, or other objects) with human-readable text describing what the names are meant to denote, and formal axioms that constrain the interpretation and well-formed use of these terms"* (Gruber 1994).

To define a common terminology of Design Thinking related terms, we created a Design Thinking ontology. With the help of the Design Thinking ontology, we describe which kind of knowledge should be included in a Design Thinking project documentation. Thus, the Design Thinking ontology defines which kind of knowledge needs to be retrievable later on and, therefore, needs to be recovered before.

We use Unified Modeling Language (UML) class models in combination with stereotypes to describe our Design Thinking ontology. The classes in the UML class model describe either artifact types or annotation types. Artifact types (cf. stereotype ArtifactType) describe which kinds of artifacts are included in a Design Thinking project documentation, while annotation types (cf. stereotype AnnotationType) describe which kind of knowledge a Design Thinking project documentation should cover. Furthermore, artifacts of certain artifact types and annotations of certain annotation types can act in terms of certain roles that are described by role types (cf. stereotype RoleType) within our Design Thinking ontology.

In Design Thinking, project documentation consists of folders and files (artifacts for short). Furthermore, approx. 90 % of all artifacts are images that are in general not machine-readable (Beyhl and Giese 2015). Therefore, meta-data such as creation dates, keywords, file names, file hierarchies, and file versions become even

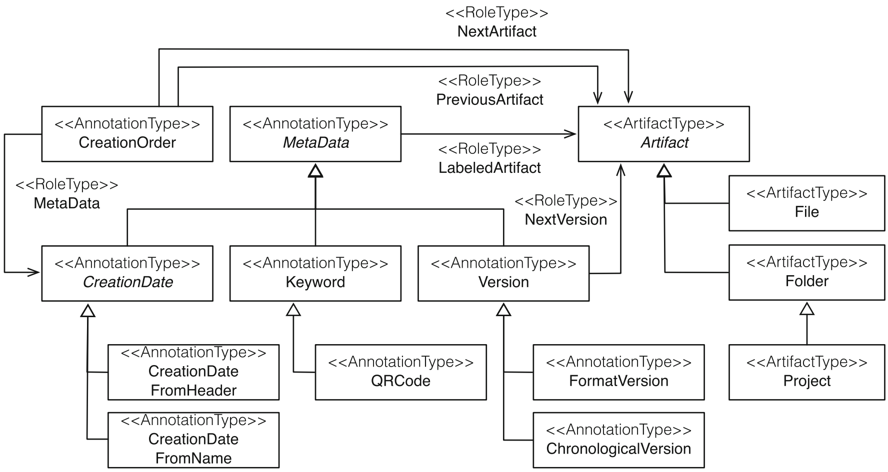


Fig. 3 UML class model: meta-data in design thinking

more important when recovering the desired knowledge. Figure 3 depicts relevant meta-data in Design Thinking project documentation that need to be extracted for further processing. As depicted, each artifact can be labeled with meta-data such as creation dates, keywords, and versions of an artifact. For example, keywords can be entities such as stakeholders or process activities. QR codes are digital markers within photographs that tag them explicitly with pre-defined keywords. Furthermore, extracted creation dates can be used to recover the chronological order of artifacts.

The extracted meta-data can be used to recover information about employed techniques, activities, and process steps. Figure 4 depicts that each artifact in a Design Thinking project documentation is the outcome of a certain phase. In Design Thinking, a methodology phase can be a process step, activity, or technique (design phase for short). A technique is the implementation of an activity. A process step can consist of multiple activities that are considered as implementation of the process step. Furthermore, a design phase can be the follow-up design phase of another design phase. A follow-up phase is different from the design phase before, e.g. a test phase is the follow-up phase of a prototype phase. Further, we distinguish follow-up phases into continuations and iterations. In contrast to a follow-up design phase, a continuation describes the proceeding of a design phase, e.g. the continuation of a test phase. An iteration describes the repetition of a certain design phase, e.g. 2nd prototype iteration.

Figure 5 depicts the ontology of design artifacts. These design artifacts are of special interest, because it is very likely that these artifacts embody the design rationales that justify next design activities. We distinguish between process artifacts, inspirational artifacts, and milestone artifacts. Process artifacts are artifacts that steer the overall process such as research questions, personas, synthesis documents, how-might-we questions, selected ideas, and prototypes. Transition

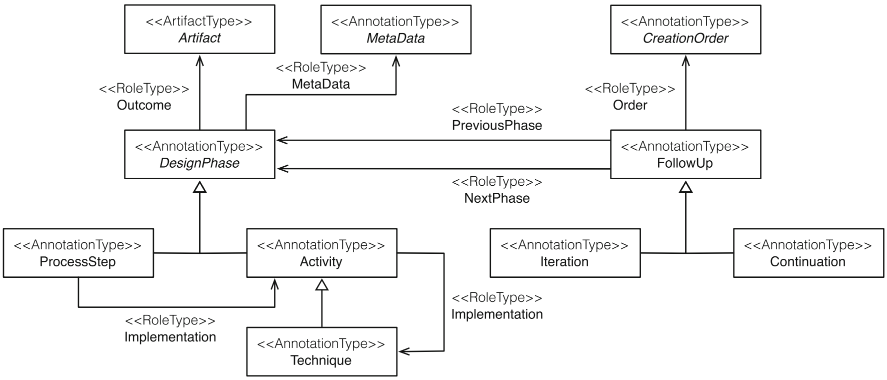


Fig. 4 UML class model: methodology in design thinking

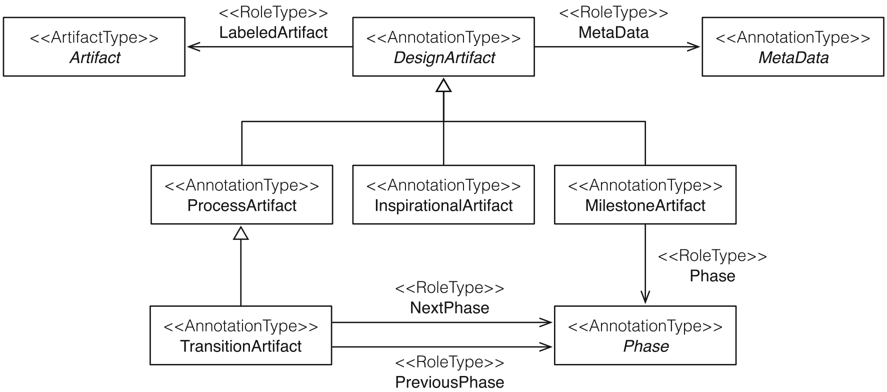


Fig. 5 UML class model: design artifacts in design thinking

artifacts are special process artifacts that mark the transition from one design phase to another design phase. Transition artifacts are of special interest, because it is very likely that they embody design rationales that justify the follow-up design phase by summarizing the working result of the previous design phase. Milestone artifacts are artifacts that mark milestones in the overall process such as project reports, intermediate documentation (e.g. LogCal, Menning et al. 2014), and final documentation (e.g. LogBook). Milestone artifacts summarize the project progress and are often the artifacts that are often created at the very end of a certain design path. Inspirational artifacts are artifacts that are related to the design challenge and inspire the design team such as souvenirs.

Figure 6 depicts involved stakeholders. Stakeholders are the design team itself, teachers in educational settings, project partners, and prospective end users. End users can be extreme users. The knowledge about stakeholders is important to differentiate findings and feedback, since it makes a difference whether the feedback was given by a prospective end user or a project partner.

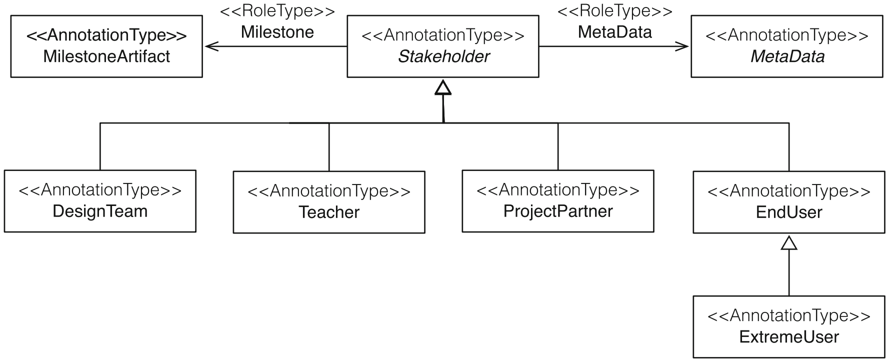


Fig. 6 UML class model: stakeholders in design thinking

3.2 Recovery Model

Figure 7 depicts our Design Thinking design path recovery model, which consists of multiple related recovery modules. In this section, we introduce some of our Design Thinking recovery modules and outline their purposes. Furthermore, we describe the interplay of the recovery modules.

We employ a naming scheme for recovery modules, e.g. R01a. Recovery modules that recover the same kind of knowledge consist of the same number (e.g. “01”). Recovery modules that recover the same kind of knowledge by employing different procedures consist of the same number, but different suffix letters (e.g. “a”) to be differentiable.

The recovery modules without ingoing edges are considered as low-level recovery modules. The purpose of low-level recovery modules is to extract relevant meta-data from artifacts and provide the extracted meta-data to dependent recovery modules. For example, file names and file headers are used to extract such meta-data as summarized in Table 1.

High-level recovery modules use the extracted meta-data. For example, recovery modules that reconstruct the employed methodology analyze the extracted keywords for each artifact to infer in which design phase the artifact was created. Further, recovered knowledge about low-level methodology knowledge such as employed techniques can be used to recover high-level methodology knowledge such as employed activities and process steps. Table 2 gives an overview about the recovery modules for the reconstruction of the employed methodology.

Further, the recovered knowledge about the employed methodology can be used to identify process artifacts and milestone artifacts. For example, the artifact created at the very end of a certain design phase (e.g. point of view) is often a milestone artifact (e.g. persona), because it summarizes the working state. Table 3 summarizes the recovery modules for the detection of design artifacts.

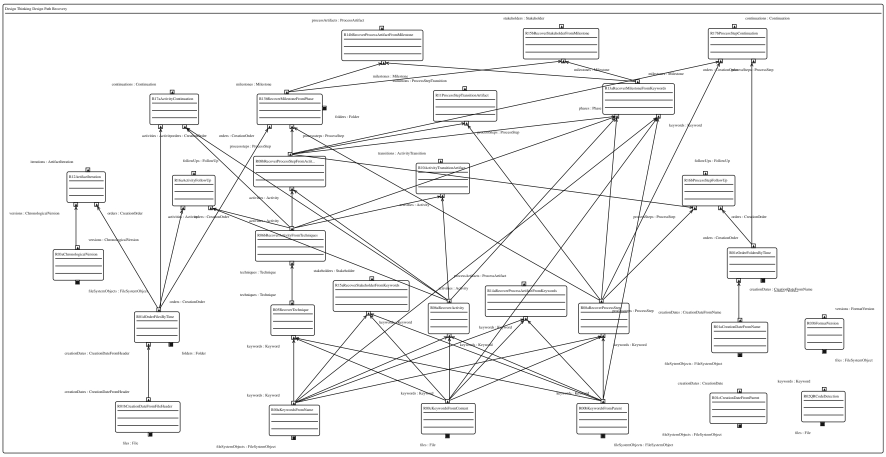


Fig. 7 Rule catalog for Design Thinking design path recovery

Table 1 Recovery modules for the extraction of meta-data

Recovery module no.		Purpose	Exploited data (depends on)
R00	a	Keyword extraction	File name
	b		Parent's file name
	c		File content
R01	a	Creation date extraction	Date encoded in file name
	b		Date stored in file header
	c		Date encoded in parent's file name
	d	Extraction of creation order	Folders in file system and creation dates extracted from file headers (cf. R01b)
	e		Folders in file system and creation dates extracted from file names (cf. R01a)
R02		Keyword extraction	Explicit QR codes in images
R03	a	Extraction of chronological artifact version	Files in same folder
	b	Extraction of artifact version stored in different format	Files in same folder

Milestone and process artifact knowledge can be used to identify which artifact embodies knowledge about stakeholders, because for certain kinds of artifacts it is rather likely that they embody knowledge about stakeholders than for other artifacts. For example, a persona embodies knowledge about a fictional stakeholder that describes the needs of the prospective end-user of an innovation. Table 4 summarizes the recovery modules for recovering stakeholder knowledge.

Table 2 Recovery modules for the reconstruction of the employed methodology

Recovery module no.		Purpose	Exploited data (depends on)
R05		Recover employed technique	Extracted keywords (cf. R00)
R06	a	Recover employed activity	Extracted keywords (cf. R00)
	b		Recovered techniques (cf. R05)
R08	a	Recover employed process step	Extracted keywords (cf. R00)
	b		Recovered activities (cf. R06a and R06b)
R16	a	Recover follow up activities	Creation order of artifacts (cf. R01d) and recovered activities (cf. R06)
	b	Recover follow up process steps	Creation order of folders (cf. R01e) and recovered process steps (cf. R08)
R17	a	Recover continuations of activities	Creation order of artifacts (cf. R01d) and recovered activities (cf. R06)
	b	Recover continuations of process steps	Creation order of folders (cf. R01e) and recovered process steps (cf. R08)

Table 3 Recovery modules for the detection of milestone and process artifact

Recovery module no.		Purpose	Exploited data (depends on)
R10		Recover transition artifact between activities	Recovered activities (cf. R06a and R06b)
R11		Recover transition artifact between process steps	Recovered process steps (cf. R08a and R08b)
R13	a	Identify milestone artifacts	Extracted keywords (cf. R00) and recovered design phase knowledge (cf. R06 and R08)
	b		Recovered creation order of artifacts (R01d) and recovered process steps (cf. R08)
R14	a	Identify process artifacts	Extracted keywords (cf. R00)
	b		Recovered milestone artifacts (cf. R13)

Table 4 Recovery modules for the identification of artifacts that embody stakeholder knowledge

Recovery module no.		Purpose	Exploited data (depends on)
R15	a	Identify artifacts that embody stakeholder knowledge	Extracted keywords (cf. R00)
	b		Recovered milestone artifacts (cf. R13)

3.3 Recovery Modules

In this section, we describe how our Design Thinking ontology is instantiated with the help of recovery modules by discussing some recovery module examples. Recovery modules implement knowledge extraction and recovery algorithms to reconstruct the knowledge as defined by the Design Thinking ontology. Note, that different procedures can be employed to recover the same kind of knowledge. Therefore, multiple recovery modules can exist that recover the same kind of knowledge, but employ different procedures to do so. However, at least one recover module per annotation type in the Design Thinking ontology should exist.

In the following paragraphs, we describe the main ideas how the recovery modules reconstruct certain kind of knowledge. According to our Design Thinking ontology, we distinguish between different categories of recovery modules, e.g. recovery modules that extract meta-data for further processing, recovery modules that reconstruct the methodology at work such as knowledge about employed process steps, activities, and techniques, and recovery modules that provide an approximation which design artifacts contain design rationales and design alternatives.

We provide a default pattern language for implementing recovery modules as depicted by Fig. 8. Solid rectangles denote Design Thinking artifacts of a certain artifact type. Dashed rounded rectangles denote annotations of a certain annotation type, i.e. recovered knowledge, that are created when the pattern and additional constraints are fulfilled. Solid edges denote references between Design Thinking artifacts. Dashed edges denote roles in which certain artifacts or annotations participate in an annotation.

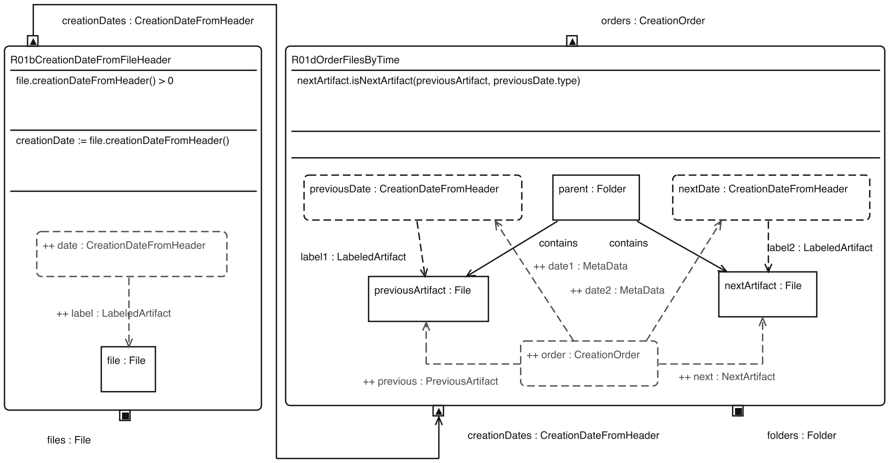


Fig. 8 Recovery modules R01b and R01d for the extraction of creation dates from file headers and the reconstruction of the temporal order of design artifacts

First, meta-data embodied within design artifacts need to be extracted to enable dependent recovery modules to exploit these meta-data for their processing. The most relevant meta-data contained by design artifacts are creation dates, file names, and file content in case the design artifact is machine-readable. Furthermore, the same kinds of meta-data are inherited from parent artifacts such as folders to child artifacts such as files. Therefore, the file hierarchy is also an important meta-data that should be exploited during recovery and, therefore, should be made available to dependent recovery modules.

According to our Design Thinking ontology, in Fig. 8 creation dates of design artifacts can be either extracted from the file header, file name, or the name of the file's parent artifact. For example, Fig. 8 depicts the recovery module R01b on the left hand side that extracts creation dates from file headers and provides the extracted creation date meta-data (cf. *date* annotation) via the output connector *creationDates* to dependent recovery modules. Such a dependent recovery module is the recovery module R01d on the right hand side of Fig. 8. The recovery module R01d reconstructs the temporal order (cf. *order* annotation) of design artifacts (cf. *previousArtifact* and *nextArtifact* artifact) within a common *parent* folder when the Object Constraint Language (OCL) expression on top evaluates to true. This temporal ordering is important to recover the design path in correct chronological order.

Further, keywords can be extracted from textual content such as file names, the file content, and the name of the file's parent artifact. Exemplarily, Fig. 9 depicts the recovery module R00a for the extraction of keywords from file and folder names. The recovery module processes file system objects (i.e. files, folders, projects) as defined by the artifact connector *fileSystemObjects* at the bottom of the recovery module and produces knowledge about keywords as defined by the output connector *keywords* on top of the recovery module. The implementation of the recovery module shows that for each file system object named *fso* the file name is processed

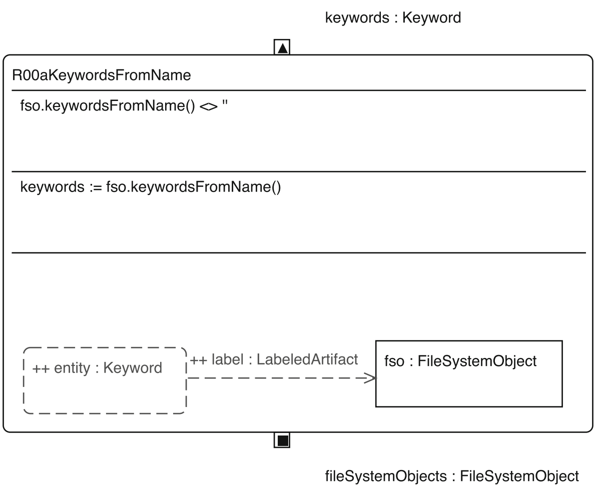


Fig. 9 Recovery module R00a for the extraction of keywords from names of file system objects

by the function *keywordsFromName()* to extract keywords. If the function *keywordsFromName()* returns a non-empty string of keywords, a *keyword* annotation is attached to the processed file system object. The *keyword* annotation acts in terms of a *label* that is attached to the file system object named *fso*. Furthermore, the extracted keywords are added as *keywords* attribute to the *keyword* annotation. Finally, all created keyword annotations are provided to dependent recovery modules via the output connector *keywords*. Other meta-data can be extracted analogously.

For example, in case the files under investigation are images, machine-readable quick response (QR) codes with special meanings can be identified within these images. Such QR codes are some kind of special keywords that are explicitly attached to whiteboards before taking photographs of these whiteboards. Versions of artifacts are also an important meta-data, because they may indicate design iterations or design alternatives. We distinguish between chronological design artifact versions and artifact versions in different file formats. Artifact versions in different file formats enable to inherit recovered properties between artifact versions, because different file formats carry different kind of meta-data. In contrast, chronological versions of design artifacts indicate design iterations or design alternatives. Chronological versions of files can be identified by file name suffixes, e.g. v1, v2, v3.

The extracted meta-data can be used to recover knowledge about the employed methodology at work. According to our Design Thinking ontology, knowledge about the methodology at work includes knowledge about employed process steps, activities and techniques. As described by our Design Thinking ontology, each design artifact is the outcome of a certain design phase. Already extracted keywords can be analyzed concerning indicators that enable to recover in which design phase a certain design artifact was created.

For example, Fig. 10 depicts the recovery module R05 that recovers the technique, which was used to create a certain design artifact. The recovery module R05 requires knowledge about keywords as defined by the input connector *keywords* at the bottom of the recovery module and produces knowledge about employed techniques as defined by the output connector *techniques* on top of the recovery module. The implementation of the recovery module shows that *keyword* annotations are analyzed by the function *technique()*. If the function *technique()* returns a non-empty string, then a concrete technique could be recovered and a *technique* annotation is attached to the exploited *keyword* annotation and the design *artifact* that is the *outcome* of the employed technique. Furthermore, the name of the technique is attached as *name* attribute to the *technique* annotation. Finally, all recovered techniques are provided via the output connector *techniques* to dependent recovery modules. Knowledge about employed activities and process steps can be recovered analogously.

Employed techniques often indicate which activity has been employed. This is possible, because certain techniques are characteristic for certain activities. For example, storytelling is a characteristic technique for the unpacking activity. Therefore, knowledge about employed techniques can be exploited to recover

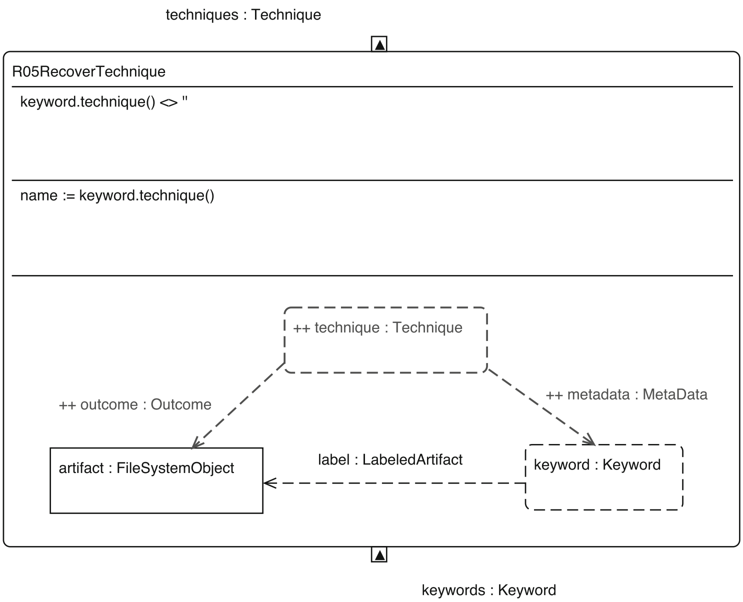


Fig. 10 Recovery module R05 for the recovery of employed techniques

employed activities. Figure 11 depicts the recovery module R06b that recovers activities based on employed techniques. The recovery module takes knowledge about employed techniques as defined by the input connector *techniques* and provides recovered knowledge about activities via the output connector *activities*. The function *activityFromTechnique()* checks which activity is the most likely one for the *technique* that was used to create the *artifact*. In case an activity could be found, the activity name is assigned to the *activity* annotation as *name* attribute. Finally, the recovered activities are provided via the output connector *activities*. Analogously, knowledge about employed activities can be exploited to recover employed process steps, because certain activities are characteristic for certain process steps.

The recovered activities and process steps can be used to identify artifacts that constitute the transition between two subsequent activities respectively process steps. These artifacts are especially important, because it is very likely that they embody important design rationales, which justify the next activity respectively process steps.

Figure 12 depicts the recovery module R11 that identifies artifacts, which mark the transition between two subsequent process steps. The recovery module requires knowledge about process steps as defined by the input connector *processSteps* and provides knowledge about transition artifacts via the output connector *transitions*. The implementation of the recovery modules shows that an artifact is considered as transition artifact when the artifact consists of two process step annotations that describe different process steps (cf. value comparison of *name* attributes in terms of an OCL expression on top of the recovery module). In case the condition is fulfilled, a *transition* annotation is created that references both process step annotations

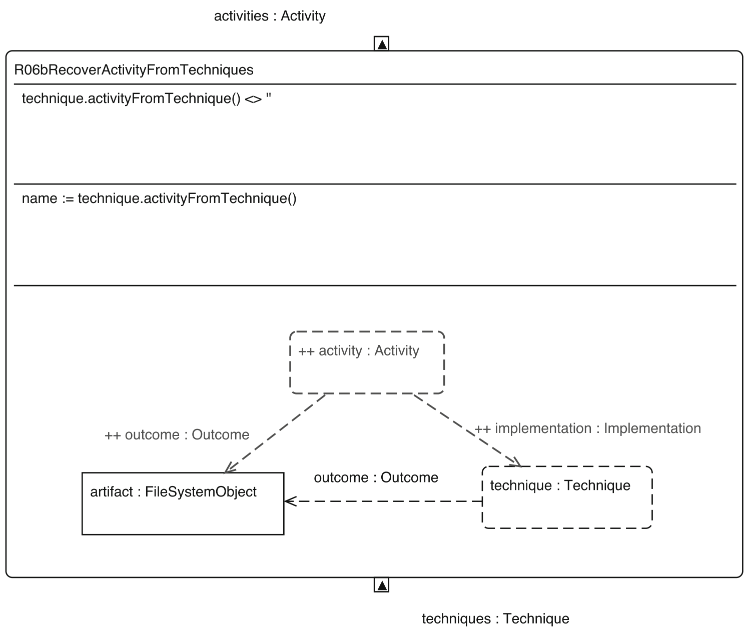


Fig. 11 Recovery module R06b for the recovery of employed activities based on recovered techniques

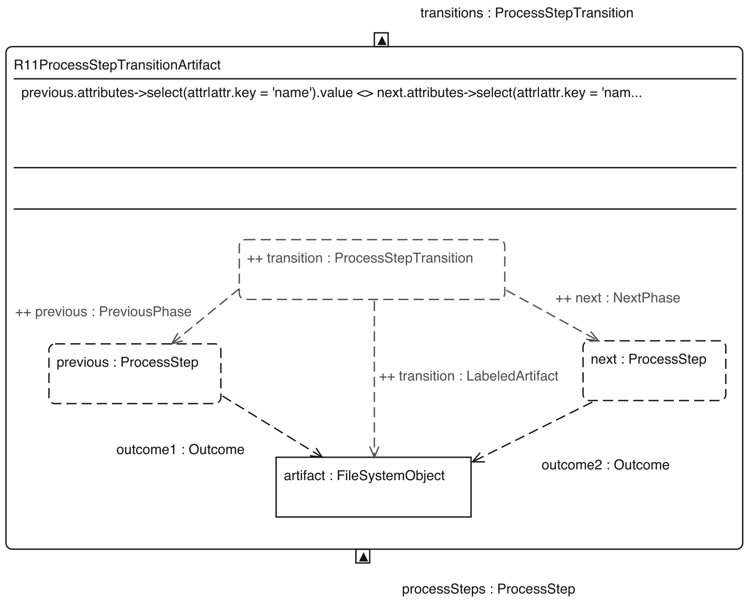


Fig. 12 Recovery module R11 for the recovery of artifacts that constitute the transition between two subsequent process steps

(cf. *previous* annotation and *next* annotation) and the recovered transition *artifact*. Analogously, knowledge about artifacts that mark the transition between activities can be recovered in the same manner.

Besides artifacts that mark the transition from one design phase to another design phase, so called milestone artifacts exist that summarize the working state at several points in time during the design process. These milestone artifacts are worth to be recovered, because it is very likely that they embody the justification for taken design decisions and describe the outcome of a design phase. Several approaches are imaginable how to identify such milestone artifacts within Design Thinking project documentation. The straightforward approach is to lookup milestone artifacts with the help of common keywords used for naming milestone artifacts. For example, in educational Design Thinking settings special intermediate reports and presentations are kindly requested by teachers and project partners.

For example, the LogCal (Menning et al. 2014) is used for daily documentation, the LogBook is used for final documentation, and intermediate presentations are given to project partners and other design teams to share ideas and receive feedback. A more sophisticated approach to identify milestone artifacts bases on the assumption that artifacts created at the very end of a certain design phase are milestone artifacts. Therefore, knowledge about the chronological order of artifacts and design phases need to be combined to identify milestone artifacts in Design Thinking project documentation.

Figure 13 depicts the recovery module R13b that identifies milestone artifacts within a certain process step. Since knowledge about process steps and creation orders is required to recover milestone artifacts, the recovery module consists of input connectors that consume knowledge about *process steps* and creation *orders*. Further, folders are often used to organize files that belong to the same process step. We consider the parent folder as container for artifacts that belong to the same process step. Therefore, the recovery module R13b also consumes knowledge about folders via the artifact connector *folders*. As depicted by Fig. 13, the last created artifact within a process step is the artifact that does not act in a role of a previous artifact in creation *order* annotations. Figure 13 depicts this negative application condition in terms of a crossed out creation *order* annotation. Finally, recovered milestone artifacts are provided via the *milestones* output connector to dependent recovery modules.

Process artifacts are artifacts that steer the overall design process. For example, a persona summarizes the needs of a fictional prospective end-user, while how-to-questions serve as basis for ideation. Besides exploiting extracted keywords to identify such process artifacts, recovered milestone artifacts can be exploited to identify process artifacts. For example, a persona is often the result and last created artifact of the point of view process step and also an important process artifact, because it serves as basis to create how-to questions for ideation.

Figure 14 depicts the recovery module R14b that recovers process artifacts based on knowledge about milestone artifacts. The implementation of the recovery module R14b shows that milestone artifacts and their attached phase annotation are used to map milestone artifacts to process artifacts (cf. OCL expression on top of the recovery module).

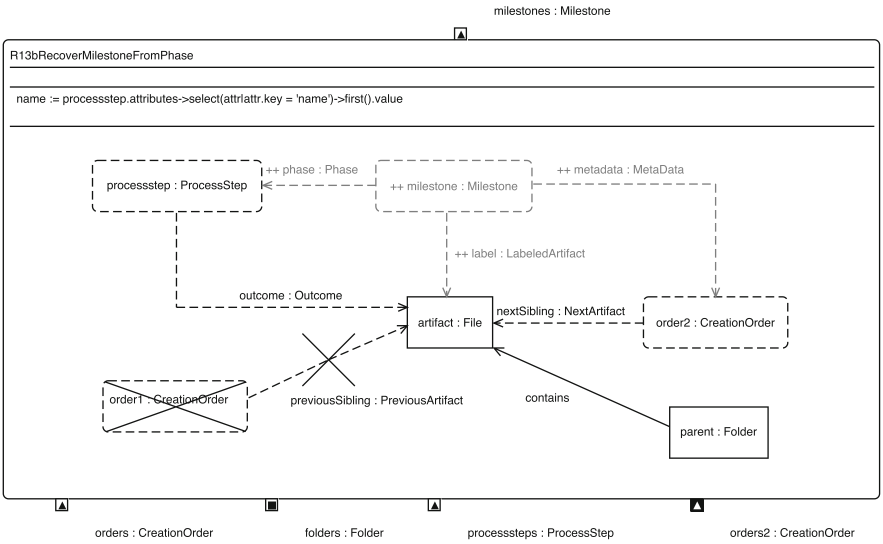


Fig. 13 Recovery module R13b for the recovery of milestone artifacts

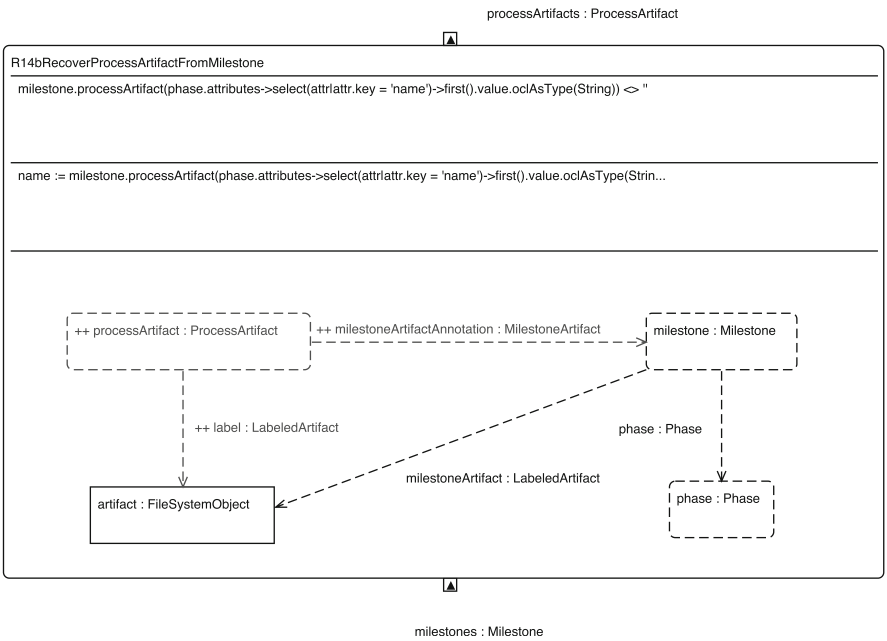


Fig. 14 Recovery module R14b for the recovery of process artifacts

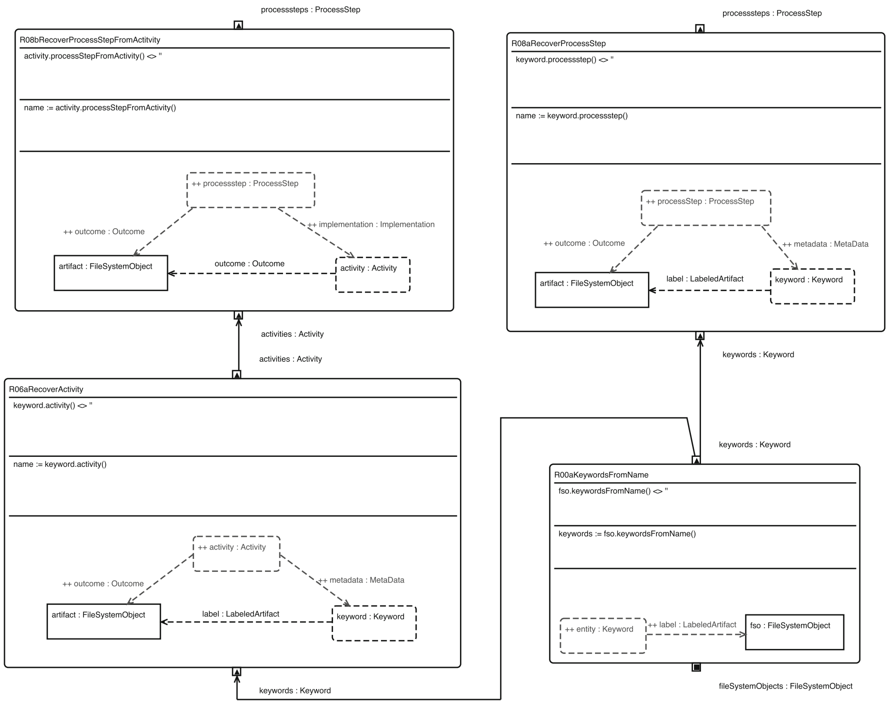


Fig. 15 Recovery module R08a and R08b that recover knowledge about employed process steps based on knowledge about keywords respectively employed activities

Figure 15 depicts how different knowledge sources are exploited to recover knowledge about employed process steps. On the right hand side of Fig. 15, the recovery module R00a provides knowledge about extracted keywords that is used by the recovery module R08a to infer the process step in which the investigated design artifact was created. On the left hand side of Fig. 15, the recovery module R06a exploits the knowledge about extracted keywords as well to recover knowledge about employed activities that lead to the artifact under investigation. Furthermore, the knowledge about employed activities created by recovery module R06a is used by recovery module R08b to infer the process step in which the investigated artifact was created.

4 Evaluation

In this section, we give an evaluation of our Design Thinking recovery model. For that, we distinguish between a qualitative evaluation (Sect. 4.1) and a quantitative evaluation (Sect. 4.2). In the qualitative evaluation, we provide recovery examples and argue why the recovered knowledge must be correct. In the quantitative evaluation, a student assistant rates the recovered knowledge from his perspective in a subjective manner. Finally, we discuss the validity of our evaluation (Sect. 4.3).

4.1 Qualitative Evaluation

In this section, we show multiple recovery examples. For that, we took a well-organized Design Thinking project documentation from an educational Design Thinking project with a duration of 12 working days and applied our Design Thinking recovery model.

In the following figures, solid rectangles denote artifacts such as file and folders. Solid lines between solid rectangles denote that an artifact contains another artifact. The filled rhombus denotes the super ordinate artifact. Dashed rounded rectangles denote annotations, i.e. recovered knowledge. Dashed lines denote roles of artifacts respectively annotations.

Figure 16 depicts an example how the storytelling folder was assigned to the process step point of view. The recovery module R05 recovered that the storytelling folder is the result of a storytelling activity, because the folder is named "storytelling". Further, the recovery module R06b recovered that the storytelling folder belongs to an unpacking activity, because the storytelling technique is often used to unpack observations and gained insights. Since the storytelling folder contains photographs with images of clustered observations, it is very likely that the unpacking activity was inferred correctly. Next, the recovery module R08b recovered that the storytelling folder belongs to the process step point of view, because the unpacking activity is a common activity in the point of view phase. Since the

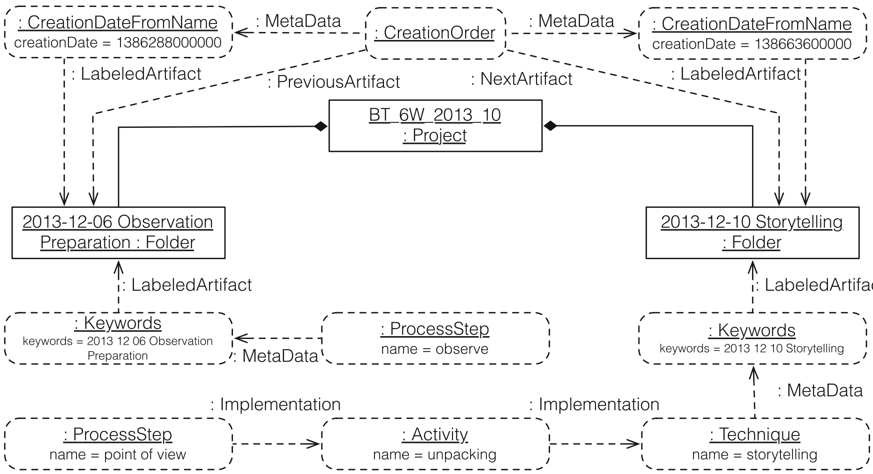


Fig. 16 Recovery of process step knowledge based on knowledge about employed activities and techniques

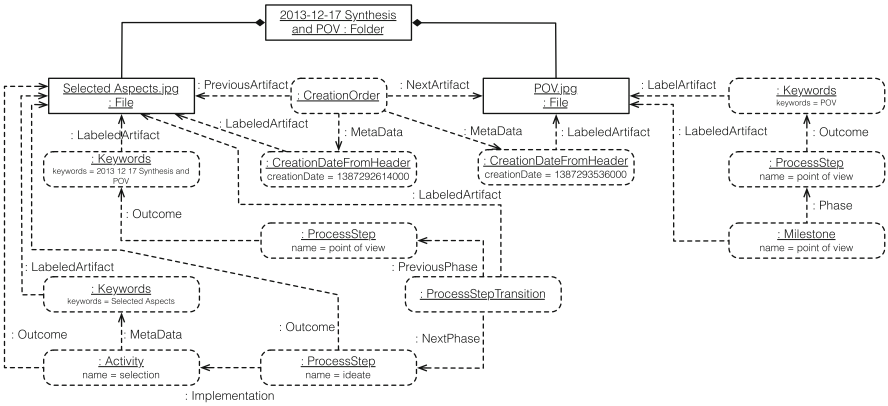


Fig. 17 Recovery of process step transitions and milestones based on process step knowledge

previously created folder is named "observation preparation" and, therefore, indicates the process step observe, it is very likely that the storytelling folder is assigned correctly to the subsequent process step "point of view". Note, since the storytelling folder is assigned to the point of view process step, also all files contained by the storytelling folder are assigned to the point of view process step as well.

Figure 17 depicts an example with two photographs contained by a folder named "Synthesis and POV". Further, the photograph "selected aspects.jpg" was created before the photograph "POV.jpg" as denoted by the creation order annotation between both artifacts. Thus, the point of view (POV for short) was derived from certain selected aspects. Note, that the recovery module R00b inherited the keywords extracted for the folder to the contained files. For example, the photograph "selected aspects.jpg" is labeled with the keywords "Synthesis and POV" and, therefore, is assigned to the process step "point of view" by the recovery module R08a. Additionally, the recovery module R00a used the name of the photograph "selected aspects.jpg" to extract the additional keywords "selected aspects". Since the word stem "select" indicates a selection activity, the recovery module R06a assigned the "selected aspects.jpg" photograph to the activity "selection", which in turn was exploited by the recovery module R08a to assign the photograph to the process step "ideate". The recovery module R08a assigned the photograph to the process step "ideate", because selection activities are common during the ideate process step. For example, it is selected which need is addressed by a subsequent brainstorming activity or which ideas are taken to the prototype process step. Since the "selected aspects.jpg" artifact is assigned to two process steps, the recovery module R11 inferred that the "selected aspects.jpg" photograph is an artifact that marks the transition between the process step point of view and the process step ideate. This piece of recovered knowledge is correct, because the "selected aspects. jpg" summarizes three selected needs used for ideation later on.

Moreover, the "POV.jpg" artifact was assigned to the process step "point of view" by recovery module R08a due to its name. Since the "POV.jpg" artifact is the

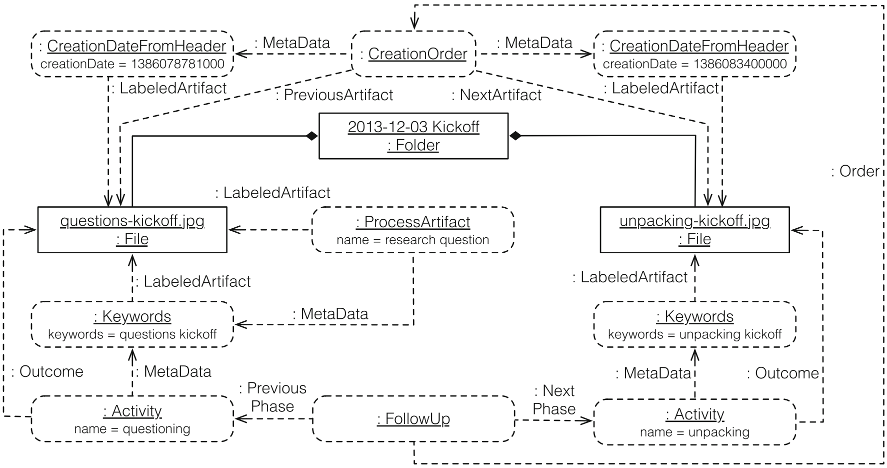


Fig. 18 Recovery of follow-up activities based on knowledge about creation orders of artifacts and activities

last created artifact in the folder "Synthesis and POV" respectively point of view process step, the "POV.jpg" artifact is identified as milestone artifact by recovery module R13b. This piece of recovered knowledge is correct, because the "POV. jpg" artifact embodies a photograph of a persona, which summarizes the needs of a fictional person that is used for ideation later on.

Figure 18 depicts that the unpacking activity is a follow-up activity of the questioning activity. The recovery module R00a extracted the keywords for the artifact "questions-kickoff.jpg" and "unpacking-kickoff.jpg". Next, the recovery module R06a assigned the artifacts to the questioning activity respectively unpacking activity. Since the "questions-kickoff.jpg" artifact was created before the "unpacking-kickoff.jpg" artifact as derived by recovery module R01d, the questioning activity happened before the unpacking activity. Thus, the unpacking activity is a follow-up activity of the questioning activity. This piece of knowledge is correct, because the "unpacking-kickoff.jpg" artifact shows a mind-map that answers some of the previously stated questions. Furthermore, the "questions-kickoff.jpg" artifact was identified as process artifact that contains research questions by recovery module R14a. This piece of knowledge is correct, because the "questions-kickoff.jpg" artifact embodies a photograph of post-its that state several questions at the beginning of the project.

4.2 Quantitative Evaluation

In our quantitative evaluation, we asked a student assistant to evaluate all created annotations, i.e. recovered knowledge, concerning their correctness on a Likert

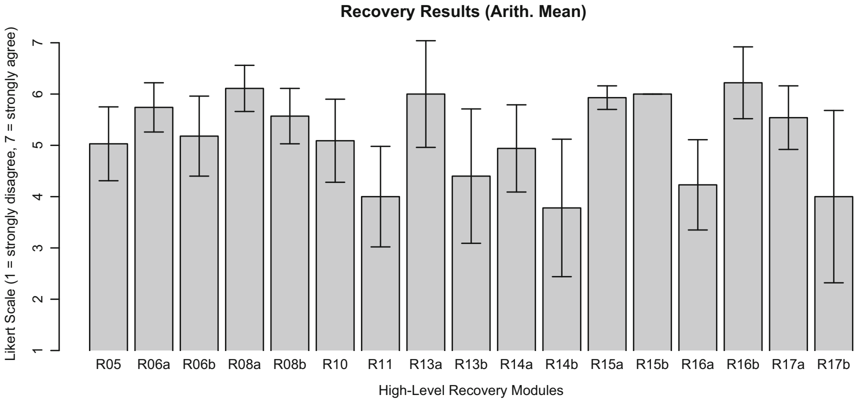


Fig. 19 Arithmetic mean of Likert values and confidence intervals per recovery modules

scale independently from the researchers, who implemented the recovery modules. We have chosen to employ a one to seven Likert scale. The Likert value of one denotes a strong disagreement, while the Likert value of seven denotes a strong agreement.

For our quantitative evaluation, we took a well-documented Design Thinking project documentation in terms of a file system structure. The project documentation consists of overall 166 design artifacts that are stored in a single-level file system structure. Note, that the student assistant only evaluated created annotations. Therefore, we are only able to speak about true positives and false positives in the following paragraphs.

Figure 19 depicts a bar chart that shows the evaluation results for our high-level recovery modules. We neglected low-level recovery modules, i.e. recovery modules that only extract meta-data, in our evaluation, because they just extract meta-data and, therefore, always yield correct results. Each bar of the bar chart denotes the arithmetic mean of all Likert values of all annotations created by a certain recovery module. Furthermore, the confidence interval is depicted as thin line on top of each bar. Note, that the confidence interval is computed over all created annotations per module and defines the Likert value range in which 95 % of all annotations are situated concerning their correctness. As shown by Fig. 18, the recovery results are in general promising, because the arithmetic mean is in general above the Likert value of 4.

Especially, the recovery modules that recover the employed techniques (R05), activities (R06), and process steps (R08) yield results that are agreed by the student assistant with a Likert value between 5.03 and 6.11. Based on these results, the recovery module that detects transition artifacts between activities (R10) yields positive recovery results with a Likert value of 5.09. The recovery modules that detect transition artifacts between process steps (R11) yield neutral recovery results with a Likert value of 4.00. Furthermore, the knowledge about employed process steps and activities is exploited by the recovery modules R16 and R17 to reconstruct follow-up activities and process steps, respectively, continuations of activities and process steps. While the recovery of follow up activities (R16a) yields neutral results with a Likert value of 4.23, the recovery of follow up process steps (R16b) yields good results with a Likert value of 6.22. In turn, the recovery of design phase continuations yields positive results for continued activities (R17a) with a Likert value of 5.54 and neutral results for process steps (R17b) with a Likert value of 4.00.

The recovery modules that identify milestone artifacts (R13) and process artifacts (R14) also yield promising results. While, the recovery module (R13a) yields milestone artifacts with a Likert value of 6.00, the recovery module (R13b) yields milestone artifacts with a Likert value of 4.40. Further, the recovery module R14a detects process artifacts with a Likert value of 4.94 and the recovery module R14b detects process artifacts with a Likert value of 3.78.

The recovery module R15 detects stakeholders within the project documentation. Both procedures (i.e. R15a and R15b) yield agreed results with a Likert value of 5.93 and 6.00, respectively.

4.3 Threats to Validity

In the above sections, we performed a qualitative as well as quantitative evaluation. On the one hand, the qualitative evaluation is subjective, but demonstrates the correctness of recovery results by example. On the other hand, the quantitative evaluation was performed by a student assistant independent of the researchers, who implemented the recovery modules. Therefore, the quantitative evaluation is objective and considers all created annotations in contrast to the qualitative evaluation.

However, the student assistant evaluated only created annotations and, therefore, only commented on true and false positives. Thus, true and false negatives are not included in our evaluation. For that, all design artifacts in the project documentation have to be evaluated concerning all properties (i.e. annotation types in the Design Thinking ontology) that need to be recovered. Thus, the number of data points that need to be created manually explodes (#artifacts * #properties) when evaluating true and false negatives.

Note, that only one student assistant evaluated the recovered knowledge. Therefore, the evaluation results may be not representative. Furthermore, we only applied the recovery modules to a single project documentation and, thus, we may get different evaluation results for project documentations of other Design Thinking projects. Additionally, asking the design team, who worked on the design project, to evaluate the recovered knowledge may yield different results, because the design team was part of the design process and, therefore, is aware of more detailed knowledge about the employed methodology and important design artifacts instead of persons that do not belong to the design team.

Note also, that high-level recovery modules depend on the correctness of the recovery results of low-level recovery module. Thus, high-level recovery modules, which depend on low-level recovery modules that perform poorly, also do not yield good results. These dependencies have to be considered when assessing recovery modules.

5 Related Work

In the past several attempts have been made to capture Design Thinking artifacts for documentation purposes. For example, the Tele-Board project (Gericke et al. 2011; Gumienny et al. 2012) enables global collaboration of design teams by providing digital whiteboards with an interface similar to analog whiteboards. Further, different software tools have been developed that capture Design Thinking documentation artifacts and support their organization. For example, ConnectingInfos (Voget 2013) implements an algorithm that rates Design Thinking artifacts concerning their importance based on ingoing and outgoing connections to other artifacts. Taking the algorithm implemented within ConnectingInfos, Kroschk extended this algorithm by distinguishing between importance of artifacts and relevance of artifacts at certain points in time (Kroschk 2014). Kroschk integrated his extended algorithm into ProjectZoom (Beyhl et al. 2013c; Beyhl and Giese 2015). ProjectZoom (Beyhl et al. 2013b) is a software tool that enables the capturing of Design Thinking artifacts from different sources such as online file storages. Furthermore, ProjectZoom enables the organization of captured Design Thinking artifacts on a virtual whiteboard with the help of well-known techniques such as clustering, relations between artifacts, and highlighting of important sections within artifacts (Koch 2014).

Design Thinking is a modern form of requirements engineering. While Brown defines the three dimensions of Design Thinking (Brown 2009), i.e. desirability, viability, and feasibility, Pohl defines the three dimensions of requirements engineering (Pohl 1994), i.e. specification, agreement, and representation. In software engineering, traceability is employed to *"follow the life of requirements in both a forwards and backwards direction"* (Gotel and Finkelstein 1997). Winkler and von Pilgrim (2010) give an overview about traceability in requirements engineering.

Additional research areas, which are related to our approach, are knowledge discovery and information retrieval. Knowledge discovery deals with detecting new unknown structures in data sets (Goebel and Gruenwald 1999). Information retrieval deals with finding information according to a query stated by users (Mitra and Chaudhuri 2000). By developing recovery modules that instantiate our Design Thinking ontology, we had to identify common structures that indicate properties as defined by our Design Thinking ontology and, therefore, we deal with knowledge discovery. By instantiating the Design Thinking ontology, we enable information retrieval based on queries stated by engineers and design thinkers. At the time of writing no recovery approach for Design Thinking exists that is similar to our approach.

But, in software engineering related approaches exist that recover software engineering knowledge such as software design patterns. For example, Niere et al. (2003) describe an algorithm to recover software design patterns as described by the Gang of Four (Gamma et al. 1994). While the approach of Niere et al. is limited to software design patterns recovery, our approach is generic due to its concept of recovery modules, which enable the implementation of arbitrary and recovery domain independent recovery algorithms.

6 Conclusion and Future Work

In this chapter, we presented a recovery approach for innovation processes. The approach bases on the creation of a Design Thinking ontology and implementation of recovery modules, which instantiate this ontology in terms of a search index. This search index eases the retrieval task, because it contains important design path knowledge recovered by recovery modules. At the time of writing, no comparable recovery approach for Design Thinking exists. Existing approaches rather focus on the capturing of Design Thinking artifacts than analyzing and querying them.

In future, we are going to extend our Design Thinking ontology and set of recovery modules, respectively. We plan to introduce belief values that describe the general confidence in the recovery results of recovery modules, because different recovery procedures may have a different correctness of recovery results (i.e. precision) and different completeness of recovery results (i.e. recall). Further, we plan to propagate these belief values between recovery results to upgrade respectively downgrade the belief in high-level recovery results. These propagated belief values enable the ranking of search results. For example, recovery results with a high belief may be shown on top of the list of search results, while recovery results with a low belief may be shown at the end of the list.

Furthermore, we plan to automatically take over reliable recovery results into Design Thinking project documentation created during running design projects to support Design Thinkers in their documentation process. This may also help to improve the recovery modules itself, when Design Thinkers delete recovered information (false positives) or add missing not recovered information (false negatives).

Acknowledgement The authors thank Claudia Nicolai (D-School Potsdam) and Axel Menning (D-School Potsdam) for their input. The authors also thank their student assistants Josephine Harzmann, Manuel Hegner, Axel Kroschk, and Lukas Pirl for their support. We also give thanks to our bachelor project team *"Intelligent Organization of Design Thinking Artifacts"* for their input.

References

Arkley P, Riddle S (2005) Overcoming the traceability benefit problem. In: CORD conference proceedings, April 2005, pp 385–389

Beyhl T, Giese H (2015) Connecting designing and engineering activities II. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research—building innovators. Springer, Cham, pp 211–239

Beyhl T, Berg G, Giese H (2013a) Connecting designing and engineering activities. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research—building innovation eco-systems. Springer, Heidelberg

Beyhl T, Berg G, Giese H (2013b) Towards documentation support for educational design thinking projects. In: International conference on engineering and product design education, September 2013, pp 408–413

Beyhl T, Berg G, Giese H (2013c) Why innovation processes need to support traceability. In: Workshop on traceability in emerging forms of software engineering 2013, San Francisco, May 2013

Brown T (2009) Change by design. HarperCollins, New York, NY

Gamma E, Helm R, Johnson R, Vlissides J (1994) Design patterns—elements of reusable object-oriented software. Addison-Wesley, Reading, MA

Gericke L, Gumienny R, Meinel C (2011) Tele-Board: follow the traces of your design process history. In: Design thinking research. Springer, Berlin, July 2011, pp 15–29

Goebel M, Gruenwald L (1999) A survey of data mining and knowledge discovery software tools. ACM SIGKDD explorations newsletter

Gotel O, Finkelstein A (1997) Extended requirements traceability: results of an industrial case study. In: Proceedings of the third IEEE international symposium on requirements engineering 1997, pp 169–178

Gotel OCZ, Morris SJ (2011) Out of the labyrinth: leveraging other disciplines for requirements traceability. In: 19th IEEE international requirements engineering conference (RE), pp 121–130

Gruber TR (1994) Toward principles for the design of ontologies used for knowledge sharing? Int J Hum Comput Stud 43(5):907–928

Gumienny R, Gericke L, Wenzel M, Meinel C (2012) Tele-Board in use: applying a digital whiteboard system in different situations and setups. In: Plattner H, Meinel C, Leifer L (eds) Design thinking research—measuring performance in context. Springer, Heidelberg, pp 109–125

Koch J (2014) Visualizing and searching of design thinking project documentation. Bachelor thesis, Potsdam

Kroschk A (2014) Infering new knowledge into design thinking documentations. Bachelor thesis, Potsdam

Menning A, Beyhl T, Giese H, Weinberg U, Nicolai C (2014) Introducing the LogCal: template-based documentation support for educational design thinking projects. In: International conference on engineering and product design education

Mitra M, Chaudhuri BB (2000) Information retrieval from documents: a survey. Inf Retr 2 (2):141–163

Niere J, Wendehals L, Zündorf A (2003) An interactive and scalable approach to design pattern recovery. October 2003

Plattner H, Meinel C, Weinberg U (2009) Design Thinking—Innovation lernen—Ideenwelten öffnen. mi-Wirtschaftsbuch, Finanzbuch Verlag GmbH, March 2009

Pohl K (1994) The three dimensions of requirements engineering: a framework and its applications. In: Selected papers from the fifth international conference on advanced information systems engineering. Pergamon Press, Elmsford, NY, pp 243–258

Voget L (2013) An association-based documentation tool to support design thinking groups in using their acquired knowledge. Master thesis, June 2013

Winkler S, von Pilgrim J (2010) A survey of traceability in requirements engineering and model-driven development. Softw Syst Model 9(4):529–565