Understanding Innovation

Christoph Meinel
Larry Leifer *Editors*

Design Thinking Research

Investigating Design Team Performance

Springer

Understanding Innovation

Series Editors

Christoph Meinel, Potsdam, Germany
Larry Leifer, Stanford, USA

“Everyone loves an innovation, an idea that sells.” Few definitions of innovation are more succinct. It cuts to the core. Yet in doing so, it lays bare the reality that selling depends on factors outside the innovation envelope. The “let’s get creative” imperative does not control its own destiny. Expressed another way, in how many ways can we define innovation? A corollary lies in asking, in how many ways can the innovative enterprise be organized? For a third iteration, in how many ways can the innovation process be structured? Now we have a question worth addressing. “Understanding Innovation” is a book series designed to expose the reader to the breadth and depth of design thinking modalities in pursuit of innovations that sell. It is not our intent to give the reader a definitive protocol or paradigm. In fact, the very expectation of “one right answer” would be misguided. Instead we offer a journey of discovery, one that is radical, relevant, and rigorous.

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

Christoph Meinel · Larry Leifer
Editors

Design Thinking Research

Investigating Design Team Performance

Springer

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

Larry Leifer
Stanford University
Stanford, CA, USA

ISSN 2197-5752 ISSN 2197-5760 (electronic)
Understanding Innovation
ISBN 978-3-030-28959-1 ISBN 978-3-030-28960-7 (eBook)
https://doi.org/10.1007/978-3-030-28960-7

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

Foreword

This year marks the 20th anniversary of the Hasso Plattner Institute for Digital Engineering (HPI) in Potsdam. From the beginning, our mission has been to educate computer science students through a practical, user-centered approach based on real-life projects. Expanding on the institute's core objectives, the HPI D-School was founded in Potsdam 11 years ago. Since that time, the D-School has been teaching Design Thinking to computer science students at HPI as well as to students from other institutions around the world. Inspiration and support in establishing the D-School in Potsdam came from the Hasso Plattner Institute of Design at Stanford University in California, founded 5 years before the HPI D-School. At the Stanford d.school, Stanford students from all fields of study have the opportunity to learn Design Thinking methods for creating impact through human-centered design. At both locations, the results have been tremendous. We are pleased to report that more and more education and research facilities have set about establishing similar programs—to the great benefit of their students. Design Thinking practices have thereby been carefully adapted to the local culture. The application of the methodology has by no means been limited to campus settings. Design Thinking has been adopted and practiced by individuals and organizations in increasing numbers as a powerful framework to foster innovation in products, services, and operations, and recently in the strategy and the creation of innovation cultures.

The propagation of Design Thinking continues—despite the fact that there has been little to no understanding of how its results actually come about. With an increase in our interest and experience in Design Thinking—and its application in answering new and diverse challenges—the need to deepen our understanding of how and why Design Thinking works also grows. This is a need that fuels my ongoing interest in and support of conscientious research in the field through the Hasso Plattner Design Thinking Research Program, a research initiative conducted jointly by Stanford University in California and the Hasso Plattner Institute for Digital Engineering in Potsdam, Germany.

Since the implementation of the Design Thinking Research Program in 2008, our understanding of the methodology has increased manyfold. Researchers in North America and Europe have conducted more than 115 research projects investigating, illuminating, and making sense of Design Thinking. As a result, we have a solid body of new knowledge on the characteristics and mechanisms of effective Design Thinking tools, team dynamics, and the application of the approach in various contexts. New tools and methods, as well as training exercises based on the knowledge and insights gained in the program, build the foundation for more informed, empirically based Design Thinking practices that produce better results. One goal of the research community is to put the created insights into practical use by establishing improved content for teaching and learning Design Thinking in all HPI communities. Therefore, researchers have reviewed previous findings of the research program and build concrete practical measures that are useful to all Design Thinking practitioners, both novices and experienced design thinkers, and both teachers and students of design thinking practices.

New insights, newly developed approaches, tools and training exercises in Design Thinking need to be available to all who seek to advance innovation, whether in organizations or as individuals working to achieve social and cultural change. Extensive research conducted by the Hasso Plattner Design Thinking Research Program has yielded valuable insights concerning the workings of Design Thinking. One way of bringing our findings to innovators everywhere is through our book series *Design Thinking Research*. This series presents a comprehensive collection of research studies carried out by scholars of both the Hasso Plattner Institute in Potsdam and Stanford University. In addition to providing the findings of the most recent projects, the 10th volume of the series, which you are now holding in your hands, gives us a new perspective from which to see Design Thinking that has been applied to recent projects, namely, neuroscience methodologies and approaches. This focus offers additional ways to gain a better understanding of how Design Thinking works.

The Design Thinking Research Program has cultivated a growing community on two continents. In doing so, it has created a setting for the rich exchange between current doctoral candidates, alumni, researchers, and practitioners from diverse disciplines. Through collaboration and partnerships of many kinds, the Design Thinking Research Program brings new perspectives, insights and lasting value not only to the program and its related researchers but to Design Thinking itself and the growing community of Design Thinkers. We invite you to reach out and encourage innovators and researchers to work together, to experiment and thereby to broaden and deepen our practice and understanding of Design Thinking and how it can benefit those challenges—both large and small—facing our world today.

Palo Alto, CA, USA Hasso Plattner
Winter 2018/19

Contents

Introduction

Larry Leifer and Christoph Meinel

Abstract The Hasso Plattner Design Thinking Research Program (HPDTRP) has always put emphasis on investigating the design team in various ways. As our experience with newly emerged instruments like fMRI (functional magnetic resonance imaging) or fNIR (functional Near-Infra-Red) expands and we build on evidence from neuro-economics, neuro-marketing, neuro-engineering, and neuro-science we see a field of neuro-design evolving—a development that complements the broad Design Research landscape. This chapter invites readers to imagine how neuroscience instruments might be brought to bear on measuring and understanding design team performance better.

1 Investigation of Design Team Performance

Design Research has always sought to put the design team in the petri dish; whereas virtually all other design programs put the customer-client there. In 2011, Malte Jung delivered the thermometer (a physics metaphor) for teams in that petri dish (Jung 2011). Then, in 2012, Neeraj Sonalkar delivered an oscilloscope (an electronics metaphor) for team dynamics in the petri dish (Sonalkar 2012).

During the DTRP 2012–2013 academic year we supported our first major applications of neuroscience instrumentation, fMRI (functional magnetic resonance imaging) studies of brain structure changes brought about through creativity training (Reiss et al. 2013). We now see fMRI and its subsequent technology, fNIR (functional Near-Infra-Red) imaging of the brain during design activities, as being the physics equivalent of a microscope we can use to measure the microstructure of

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

C. Meinel
Hasso Plattner Institute for Digital Engineering, Prof.-Dr.-Helmert Straße 2-3, 14482 Potsdam, Germany
e-mail: meinel@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_1

what's going on in that petri dish. Given the thermometer, oscilloscope, and microscope we can image an age in which the design of human intelligence can keep pace with the evolution of artificial intelligence.

As our experience with these instruments expands and we build on evidence from neuro-economics, neuro-marketing, neuro-engineering, and neuro-science we see the field as evolving towards neuro-design—a development that complements the broad Design Research landscape. This evolutionary path has been visualized in Fig. 1 and we can understand it as a deep part of the evolution of "western culture".

Of course, the story we tell here is only one thread leading to neuro design. Many more stories could be told—and hopefully many more will be told, as people recognize how their own works contribute to neuro design, or endeavor to carry out novel studies to advance this emerging field.

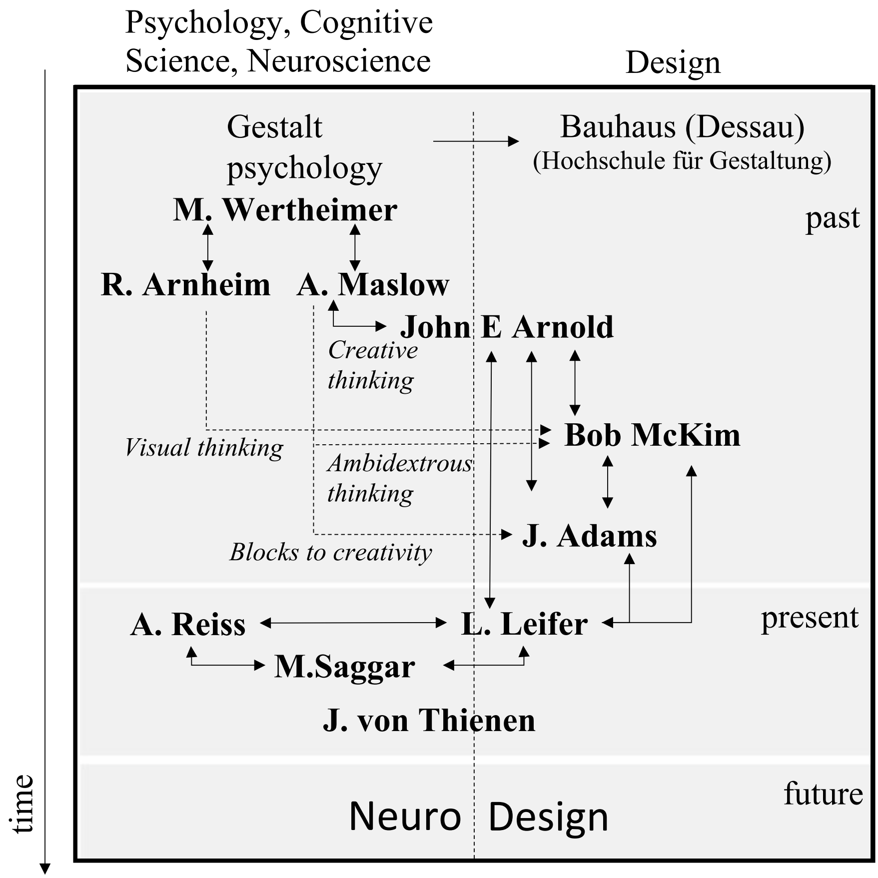


Fig. 1 The evolution of psychology, cognitive science, neuroscience, and design towards neuro design

Our story begins with Gestalt theory and the Bauhaus. This story covers a century with 2019 marking the 100th anniversary of the founding of the Bauhaus. The 2nd Bauhaus in Dessau is named “Hochschule für Gestaltung”. At the same time, there is little evidence people from Gestalt theory and Bauhaus had deeper interaction with each other. So, initially, we see rather independent roots of design on the one hand and predecessors of cognitive neuroscience on the other.

In the realm of cognition research, the legacy of Max Wertheimer is specifically noteworthy, because it has strongly impacted design thinking and recent neurodesign studies in this tradition. Wertheimer supervised both Rudolf Arnheim and Abraham Maslow, who had an influence on J. E. Arnold and his future colleagues at Stanford University. Arnold had a B.A. in Psychology before studying Engineering at MIT. Together with Robert McKim and James Adams, he started the Design Division in Stanford’s Mechanical Engineering Department. Highly integrative approaches to theorizing, research and practice in design began to take shape. In this context, cognition and design studies merged into a fruitful synthesis.

To name a few of the milestones in this story, we can appreciate how John Arnold advanced a human-centered framework of comprehensive design based on creative thinking and innovation theories (Arnold and Clancey 2016; von Thienen et al. 2017). Thus, he merged design and cognition studies, without yet including neuroscientific analyses. Robert McKim furthered the human-centric design perspective and elaborated bodily aspects in design, such as the importance of visual thinking and “ambidextrous” engagement in design activities (McKim 1972; von Thienen et al. 2018). This new synthesis included a great array of physical (“bodily”) aspects, though brain research was barely mentioned at this stage. An important question pursued over decades was how to overcome creativity blocks (Adams 1974; Arnold and Clancey 2016). Here, a lasting concern was how to help designers/creative teams use their full range of abilities, beyond rational-symbolic processing (Faste 1994; von Thienen et al. 2016). In these debates, brain research was beginning to be referenced, though not discussed in detail.

A substantial bridge between psychology/cognitive science/neuroscience on the one hand and design on the other hand was attempted by Larry Leifer, who had “insider knowledge” in both fields. After earning a master’s degree in art (industrial design), he wrote his dissertation in the Neurology department at Stanford. This was even prior to the evolution of the term “neuroscience”. Leifer’s (1969) Ph.D. research used single cell electromyography in humans performing a controlled tracking task, like driving or flying, to reveal how the nervous system was modulating force. In 1984, Leifer founded the Center for Design Research at Stanford University. A number of projects that emerged in this academic culture came to explore sweet spots between design, neuroscience and medicine (e.g., Aldaz et al. 2015; Aquino Shluzas et al. 2011; Johnson et al. 2001).

In recent years, front-end neuroscience has become an integral part of design (thinking) studies. A number of technically sophisticated research endeavours have added new details and precision to our understanding of design, creativity and innovation. At present we witness technical progress permitting the investigation of questions that were out of reach just a few years ago. Here, we can highlight the pioneering works in the laboratories of Allan Reiss, Director of the Center for Interdisciplinary Brain Sciences Research (CIBSR), and Manish Saggar, head of the Brain Dynamics Lab, at Stanford University. Their research elucidates brain dynamics, not only of individuals in laboratory settings, but of design teams working in their natural environment (Mayseless et al. 2018; Reiss 2018) using state-of-the-art technology such as Functional Near-Infrared Spectroscopy (fNIRS). Importantly, these works include the design of novel study setups and metrics that allow comprehensive investigations of creativity and design (Hawthorne et al. 2016; Saggar et al. 2015, 2017). These metrics elucidate visual processing, spontaneity, emotions and many other aspects of design performance from a neuroscientific perspective, moving far beyond more traditional studies on cognitive rational processing abilities. The historical concepts of creative, visual and ambidextrous thinking have evolved to include precise neuroscientific descriptions (measures), explanations and predictions.

Moreover, technical advancements now permit the analysis of brain dynamics of single designers on a moment-to-moment scale (Saggar et al. 2018). That is great progress compared to standard neuroimaging approaches, which average findings across several study participants and/or various study trials. Thus, it is now possible to explore patterns of individual design actions and the person's moment-to-moment neuroimaging results, opening up a vast range of new research opportunities.

Studies with this potential can be especially fruitful when they are discussed within various expert communities. Again, it is collaboration across boundaries and the union of knowledge domains (which all too often act in isolation from each other) that permit the greatest rate of learning for everyone involved. Neuroscientific research on design, creativity and innovation thrives in a network of expertise that spans diverse perspectives on these topics.

To help integrate neuroscientific perspectives and other design, creativity or innovation expertise, Julia von Thienen initiated platforms of exchange for neuroscientists and further design or creativity scholars as well as practitioners (von Thienen 2017; Reiss 2018). Her work is underpinned by cross-disciplinary reviews of study outcomes to help members of neuroscientific and other communities explore pieces of work that merge to yield the larger picture(s) of one or more overall puzzle(s) (von Thienen 2013, 2018). It was in this context that Larry Leifer came to coin the term "neuro design" to describe the happy synthesis of neuroscience and design.

Once this new field of work was discovered, many new research avenues could be explored. From a computer science perspective, Christoph Meinel noted the additional avenues of progress, besides the new study designs and formats of exchange. Obviously, a large corpus of neuroscientific data already exists. Without a doubt, many more study outcomes will emerge in the near future. What opportunities exist to learn the most from this wealth of data? Might there be opportunities for big data analyses or other technically-supported means to benefit from current research?

On the whole, we see promising examples of ground-breaking neuroscientific and design studies to further broaden our existing corpus of design research and gain new insights. Collaboration across boundaries is gaining momentum. We believe the time is ripe for even broader, orchestrated efforts to merge physiological perspectives and design research. Neuro design opens up a wide perspective for many new research questions and practical projects. It invites new unions of insight and the joining of resources for work in various directions.

As a sample topic in the broad realm of neuro design, we will briefly address the issue of creativity blocks—an important theme in design thinking traditions ever since the 1950s. Dealing with creativity blocks show us how the integration of neuroscientific perspectives can bring about the next leaps of understanding and practical implementation.

1.1 Identifying and Overcoming of Creativity Blocks

The topic of creativity blocks has been discussed in recent times by Cross (2007) in his treatise *Designerly Ways of Knowing.* Specifically, Cross points out the phenomenon of *fixation.* This concept applies when designers "may be too ready to re-use features of known existing designs, rather than to explore the problem and generate new design features" (p. 104). Typically, designers are designers of something. They own and are identified with that thing, or style of thing. To avoid the block, we prompt the designer within each of us to pause and question the problem as stated.

Does this familiar problem really deserve to be re-solved? The posing of this question stems from observations in both professional and academic environments. The constraints that define problem spaces are either shaped internally or impacted by forces extrinsic to the problem itself. In most cases these constraints prompt a quickness in the execution of design action that is perceived to be optimal, yet rarely proves to be. The quickness introduced by *solution-fixation* is for the most part not helpful, as it precludes the questioning or interrogation of problems stated by others. As a result, it prematurely freezes the problem space before it fully forms. The evolution towards neurodesign promises to afford quantitative research into the cognitive foundations of high performance design-teams and thereby to gain insight on a physical basis of how those teams overcome or even avoid creativity blocks (see Sect. 1.2). At the same time, teams or individuals might get a chance to shield themselves against those effects in situation where solution-fixation is undesirable. Thanks to the current neuroscientific studies, we are now able to identify those no longer invisible mechanisms and actively avoid practices such as creativity blocks by developing measures to guard against them.

The liabilities of *solution-fixation* are numerous; therefore our challenge is to be mindful of its causes. These causes are invariably tied to how problems themselves are first identified and managed. Design researcher Dorst (2015) names five causal syndromes of conventional problem-solving practices:

- *Lone Warrior*: A particular entity owns the problem-solution space
- *Freeze the World*: Stop the world, prevent change, and utilize static thinking
- *Self-Made Box*: Solve the problem with solutions from our past
- *Rational High Ground*: Believe in and assert our own rationality

- *Identification*: The problem and its solution are identified through organizational autopoiesis

Proceed with extreme caution when you detect these in your respective environment or culture.

Such blocks can arise both in practice and research. For instance, theories and measures of design, creativity and innovation can be primarily concerned with a person's rational abilities. This is expressed as falling victim to the "Rational High Ground" block. Neuroscientific studies recommend themselves as helpful antidotes to free us from such biases. They permit the scanning of the entire brain beyond the rather narrow bodily substrates associated with a person's rational thinking abilities. It becomes possible to elucidate how further human capacities fuel creativity, innovation and design, beyond rational processing. In this regard, notably, Saggar et al. (2017) found high creative performance associated with activity in the cerebellum (the motion coordination center in the brain) and reduced activity in the brain's pre-frontal cortex (known for rational planning).

Another level at which the block of a "Rational High Ground" can afflict research practices pertains to the scientific procedure itself. Is this research rationally preplanned all along, or does it leave room for explorations and serendipity (cf. Eris 2003, 2004): To what extend do the scientists ask deep-reasoning questions ("rational approach"), which lead to ever-deeper expertise in a single direction? To what extent do the researchers ask generative design questions (apart from rigorous rationality), which facilitate divergent thinking in science and applied fields? We envision neurodesign as a field in which deep-reasoning and generative design questions can be balanced and intermix in a synergetic interplay.

1.2 Orientation

Overcoming creativity blocks may be more or less time-consuming across domains. For instance, while problems requiring software solutions can be reasonably updated in a short amount of time, problems of the built environment require solutions with physical configurations lasting years, decades or centuries. Problems with human cognition, perception, and cultural values might take even longer to address, although neurodesign could help to speed up change.

In terms of values, design thinking has long noted the importance of mindsets. Indeed, the pursuit of any problem-oriented design action requires a human-centered mindset. In turn, problem orientation is then an accurate indicator of whether the mindset guiding action is in fact human-centered, or instead if it is rooted within any number of other centrisms. Some examples of this include systems-centrism, type-centrism, or ego-centrism.

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

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

1.3 The Applicability of Neuro Design Research

With regard to how this emerging new branch of research might complement or deepen the existing broad approaches of Design research, a number of potential connection factors have been demonstrated. But how should these extension and elaborations look like? How might we instrument design teams to measure the fixations (biases) of design teams, their problem orientation versus solution fixation? Might these instruments help us re-orient our brains? What are the structural brain differences between individuals who prefer decision-making to question-asking? Even more profoundly, might these instruments help us quantify such behaviors in collaboration—versus cooperation-scenarios (Leifer and Meinel 2018)? Might we be able to implement these behaviors on an actionable basis, day-to-day and session-to-session? Does our culture's emphasis on solutions introduce observable brain rewards that overwhelm the rewards for problem discovery? We are on the threshold of being able to implement the human brain and other physiological structures during collaboration activities.

Design Research enables us to address these questions and others like them, with new metrics and a heightened awareness of unintended biases at the pursuit of breakthrough innovations in business, government, and academia. Design Research also promises great rewards while navigating new creative possibilities by prioritizing human-centered problem formulation. For design operations to find greater resonant meaning and impact, they must feature human users and we can now instrument those humans in ways that have previously only been possible with machines. We have a profound need to better understand human designers, as artificial intelligence and robotics play an increasingly central role in the creation and delivery of new products, services, processes, and business systems. We need human intelligence evolution to keep pace with technical evolution. This is the designer's workspace of the future.

In each of the subsequent chapters we invite the reader to imagine how neuroscience instruments might be brought to bear on measuring and understanding design-team performance better.

2 The HPI-Stanford Design Thinking Research Program

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

2.1 Program Vision and Goals

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

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

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

Special interest is found in the following questions:

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

Since the start of the program in 2008 more than 100 research projects have been conducted, and our understanding of this field has advanced with the authoring of new tools and yielding of new insights. This Design Thinking Research series shares scholarly insights with a public audience, whether in a multi-national corporation or a garage-based start-up.

2.2 Road Map Through This Book

In the tenth program year of the Hasso Plattner Design Thinking research program, researchers from HPI and Stanford University have conducted a wide range of research projects on design thinking. This annual publication is a compilation of their findings, sharing outcomes arranged into four parts that illustrate the programs comprehensive approach to design thinking research.

The first part of the book is dedicated to "New Approaches to Design Thinking Education," including both analog and digitally mediated endeavors to teaching design thinking. Research-based training packages that bridge the existing gap between research and practice are proposed, as well as a tool for improving design driven creative practice in educational environments. Moreover, two ways are presented that leverage capabilities of new technologies for design thinking education: One chapter explores the use of Massive Open Online Courses (MOOCs) in design thinking learning scenarios, another chapter experiments with immersive virtual reality.

The studies clustered in part II "Exploring Effective Team Interaction" examine effective team interaction in a broad variety of ways. This interaction encompasses communication aspects, such as styles of design conversations and communication mediums for programmers and non-programmers. The examination ranges from showcasing the strength of network rotation in collective design and leveraging neuroscience to exploring team collaboration and mining the role of design reflection and associated brain dynamics.

The third part of this book introduces "Tools to Support Design Thinking Practices"—all of them leveraging digital technologies. Prototyper is a web browser-based collaborative virtual environment that supports the joint real-time creation of three-dimensional low-fidelity prototypes. Scenarios to combine AR and actuated tangibles with the potential to improve remote collaboration are proposed in the second chapter. The DT@IT Toolbox is a collection of design thinking methods targeted at design thinking novices that aims to support everyday software development activities. The last new tool presented in this edition of the DTR series is Poirot, a web inspection tool for designers that enables them to make style edits to websites using a familiar graphical interface.

The final part of this book is dedicated to "Applying Design Thinking Practices." A variety of application scenarios are showcased. First, design thinking is applied in developing the software system Tele-Board MED. The chapter captures the hands-on experience of psychotherapists when using TBM for the first time in consultation sessions with patients. In the second chapter, the design thinking methodology is applied to re-design the remote collab-spaces on the online learning platform openHPI to improve user-centeredness. The closing chapter focuses on overcoming common pitfalls of workspace (re-)design, using a theoretical perspective to reflect and shape practice.

2.3 Outlook

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

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

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

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

Acknowledgements We thank all authors for sharing their research results in this publication. Our special thanks go to Dr. Sharon Nemeth for her constant support in reviewing the contributions.

References

Adams, J. L. (1974). *Conceptual blockbusting*. Stanford, CA: Stanford Alumni Association.

Aldaz, G., Shluzas, L. A., 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. https://doi.org/10.1371/journal.pone.0121179.

Arnold, J. E., & Clancey, W. J. (2016). Creative engineering: Promoting innovation by thinking differently. Stanford Digital Repository. Available at: http://purl.stanford.edu/jb100vs5745 (Includes John Arnold's original Creative Engineering manuscript from 1959).

Aquino Shluzas, L., Steinert, M., & Leifer, L. (2011). Designing to maximize value for multiple stakeholders: A challenge to med-tech innovation. Paper presented at the International Conference on Engineering Design, ICED11, August 15–18, 2011, Technical University of Denmark, Denmark.

Cross, N. (2007). *Designerly ways of knowing (Board of International Research in Design)*. Basel: Birkhauser.

Dorst, K. (2015). *Frame innovation: Create new thinking by design*. Cambridge: MIT Press.

Eris, O. (2003). Asking generative design questions: A fundamental cognitive mechanism in design thinking. In *Proceedings of the international conference on engineering design* (pp. 19–21).

Eris, O. (2004). *Effective inquiry for innovative engineering design*. Netherlands: Springer.

Faste, R. (1994). Ambidextrous thinking. In *Innovations in mechanical engineering curricula for the 1990s*. New York: American Society of Mechanical Engineers. Retrieved January 2019 from http://www.fastefoundation.org/publications/ambidextrous_thinking.pdf.

Hawthorne, G., Saggar, M., Quintin, E.-M., Bott, N., Kienitz, E., Liou, N., et al. (2016). Designing a creativity assessment tool for the twenty-first century: Preliminary results and insights from developing a design-thinking based assessment of creative capacity. In H. Plattner, C. Meinel, & L. Leifer (eds.), *Design thinking research. Making design thinking foundational* (pp. 111–123). Cham: Springer.

Johnson, M. J., Van der Loos, H. M. H. F., Burgar, C. G., Shor, P., & Leifer, L. J. (2001). Designing a robotic stroke therapy device to motivate use of the impaired limb. In M. Mokhtari (Ed.), *Proceedings of 7th international conference rehabilitation robotics, ICORR 2001, Integration of assistive technology in the information age* (Vol. 9, pp. 123–132), Evry, France. IOS Press.

Jung, M. (2011) *Engineering team performance and emotion: Affective interaction dynamics as indicators of design team performance*. Ph.D. dissertation, August 2011.

Leifer, L. J. (1969). *Characterization of muscle fiber discharge during voluntary isometric contraction of the biceps brachii muscle in man*. Ph.D. Thesis, Stanford University, June 1969.

Leifer, L. J., & Meinel C. (2018). Introduction: Reflections on working together—Through and beyond design thining. In: In H. Plattner, C. Meinel, & L. Leifer (eds.), *Design thinking research. Making distinctions: Collaboration vs. cooperation* (pp. 1–12). Cham: Springer.

Mayseless, N., Saggar, M., Hawthorne, G., & Reiss, A. (2018). Creativity in the twenty-first century: The added benefit of training and cooperation. In H. Plattner, C. Meinel, & L. Leifer (eds.), *Design thinking research. Making distinctions: Collaboration vs. cooperation* (pp. 239–249). Cham: Springer.

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

Reiss, A. (2018). *Opening of the symposium "Neuroscience and Physiological Perspectives on Design Thinking and Creativity"*. Available at https://www.tele-task.de/series/1219.

Reiss, A., Hawthorne, G., & Saggar, M. (2013). *Impact and sustainability of creative capacity building*. HPDTRP project started 2013—proposal abstract available at: https://hpi.de/dtrp/projekte/projekte-201314/impact-and-sustainability-of-creative-capacity-building.html.

Saggar, M., Quintin, E.-M., Bott, N. T., Kienitz, E., Chien, Y.-H., Hong, D. W.-C., et al. (2017). Changes in brain activation associated with spontaneous improvization and figural creativity after design-thinking-based training: A longitudinal fMRI Study. *Cerebral Cortex, 27*(7), 3542–3552.

Saggar, M., Quintin, E.-M., Kienitz, E., Bott, N. T., Sun, Z., Hong, D. W. C., et al. (2015). Pictionary-based fMRI paradigm to study the neural correlates of spontaneous improvisation and figural creativity. *Scientific Reports, 5,* 10894. https://doi.org/10.1038/srep10894.

Saggar, M., Sporns, O., Gonzalez-Castillo, J., Bandettini, P. A., Carlsson, G., Glover, G., et al. (2018). Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature Communications, 9,* 1399. https://doi.org/10.1038/s41467-018-03664-4.

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

von Thienen, J. P. A. (2013). *Kausalniveaus*. Lengerich: Pabst Science Publishers. Available online at: https://refubium.fu-berlin.de/handle/fub188/8395?locale-attribute=en.

von Thienen, J. P. A. (2017). *Neurobiology and design thinking*. A podium discussion with Sergio Agnoli, Stefanie Faye Frank, Manish Saggar and Caroline Szymanski. Available at: http://m.tele-task.de/lecture/video/6388.

von Thienen, J. P. A. (2018). *Design thinking, the body and creativity: Exploring some bridges*. Available at: https://www.tele-task.de/lecture/video/7013.

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

von Thienen, J. P. A., Clancey, W. J., & Meinel, C. (2018). Theoretical foundations of design thinking. Part II: Robert H. McKim's need-based design theory. In H. Plattner, C. Meinel, & L. Leifer (eds.), *Design thinking research. Looking further: Design thinking beyond solution-fixation* (pp. 13–38). Cham: Springer.

von Thienen, J. P. A., Ford, C., & Meinel, C. (2016). The emergence of design thinking in Californian engineering classes: Four historic concepts worth knowing. In *Talk at the MIC conference: From creative brains to creative societies*, September 14–16, Bologna, Italy.

New Approaches to Design Thinking Education

Accessing Highly Effective Performative Patterns

Jonathan Antonio Edelman, Babajide Owoyele, Joaquin Santuber, Anne Victoria Talbot, Katrin Unger and Kira von Lewinski

Abstract Design Thinking is undergoing an exciting and critical transformation. Ad hoc content and practices, based on anecdote and experience, are being displaced by new content and practices grounded in empirical evidence and rigorous theory. To bring this new knowledge to both designers and design teams, a new approach to design instruction is required. The radical point of view of our research suggests that the work of design teams is a performative act (designing-as-performance) and that design sessions are a performance of a corpus of behaviors that constitute much of the practice of Design Thinking. Furthermore, this corpus of behaviors can be trained and learned in the form of a skills repertoire called performative patterns. Performative patterns function a shared model of action and reflection which provide structure for previously undefined content (Edelman 2019). This new approach to design education involves not only the intellectual task of designing and understanding theory but a phenomenological practice of perception-action loops between the body, the environment in which the team is situated and the artifacts-media with which the team interacts. Research-based training packages promise to provide both sound theory and highly effective performance patterns which together constitute a basis for excellence in team-based design.

J. A. Edelman (✉) · B. Owoyele · J. Santuber · A. V. Talbot · K. Unger · K. von Lewinski
Hasso Plattner Institute for Digital Engineering,
Campus Griebnitzsee, 14482 Potsdam, Germany
e-mail: jonathan.edelman@hpi.de

B. Owoyele
e-mail: Babajide.Owoyele@hpi.de

J. Santuber
e-mail: Joaquin.Santuber@hpi.de

A. V. Talbot
e-mail: avtalbot2@aol.com

K. Unger
e-mail: Katrin.Unger@hpi.de

K. von Lewinski
e-mail: kvlewi@outlook.com

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_2

For Merleau-Ponty, a human subject is not defined,
as Descartes had it, as an 'I think,' but rather as an 'I can.'
The world we experience, for Merleau-Ponty,
is a field of possibilities for skilled action.

E. Baggs and A. Chemero, The Third Sense of the Environment 2018

1 Introduction

The work of design is to see the world as a field of possibilities and the work of learning design is to develop the capacities for skilled action. Teams are the engine of the complex system of innovation, and team *performance* is a critical factor in developing new and appropriate solutions to the problems that face us as leaders and change makers. The focus of this paper is on building the foundation for developing teaching and learning materials for the cultivation of technical skills that build highly effective team interactions, which are the basis for team performance.

Design Thinking holds the promise of equipping its practitioners with the right tools and mindsets to adequately address the challenges presented in the 21st century and enabling them to implement impactful solutions. Iterations, learning from feedback and mistakes and constant improvements have become the mantra for Design Thinking (DT) practitioners.

In order "to understand why and how the Design Thinking method works on a scientific basis" (Plattner et al. 2011b) Hasso Plattner started a research program 10 years ago. The resulting body of research sheds light on Design Thinking from a variety of perspectives and aims at contributing to Design Thinking's academic advancement and ongoing discourse.

Those findings in addition to other valuable insights from related disciplines provide the perfect basis for the iteration and improvement of both DT theory, practice, and education. The novelty of the discipline itself calls for further refinement and development of both its practical implication as well as its body of theory.

In the DT community, the notions and understandings of how DT works vary greatly. One common point of view is David Kelly's framing of DT as a somehow magical process: "…we can put together a seemingly random team of designers out of who is available in the firm at that time, and in the end, magic happens: breakthrough ideas and happy clients" (Kelley 2018). Another perspective presents DT as a more structured and understandable process: "(it) can be well structured, and things that occur during that period are both repeatable and comprehensible" (Kolko 2010). "It is only the lack of understandable documentation, or the decision to not share that documentation, that creates the sense of magic" (Kolko 2015).

Sonalkar et al. describe design team performance as "a complex phenomenon that involves person, behavior and environment parameters interacting with and influencing each other over time" (Sonalkar et al. 2018). This point of view emphasizes the complexity of DT which needs to be looked at from many different scientific perspectives while integrating insights for a common understanding. It is not magic at work, but a network of various factors which can be analyzed and rigorously understood.

2 The State of Play in Design Thinking Education

Design Thinking has become a frequently used method to produce creative outcomes in different contexts. It is applied globally in a variety of various settings and formats.

Current DT training varies in scope and depth from one to three days of Design Thinking introductory formats, to extended offerings providing certificates, and programs over several academic study terms. In addition, there is increasing supply and demand of "online" DT formats for educating DT, ranging from online training within corporations (e.g., SAP and McKinsey), to Massive Open Online Courses (e.g., Design Thinking for Innovative Problem Solving by Darden School of Business) (Plattner et al. 2011a, 2012a, b, 2018; Thienen et al. 2018; Johansson-Sköldberg et al. 2013).

2.1 Design Thinking Learning Outcomes

In their paper "An educational perspective on design thinking learning outcomes," (Taheri et al. 2016) Taheri and her colleagues investigated current Design Thinking education through the lens of an educational model of learning outcomes. Taheri suggests three primary domains of Design Thinking learning outcomes, Affective Outcomes, Cognitive Outcomes, and Skill-Based Outcomes[1] all based on previous work by Bloom (1987), Gagné (1984) (see Fig. 1).

Taheri and her colleagues further argued that there was "a strong emphasis in the literature on the affective outcomes of design thinking, such as creative confidence, and the cognitive outcomes, such as mind-shifts," rather than skills.

Figure 2 shows a conceptual model from Taheri and colleagues illustrating the observed patterns of outcomes in three DT training formats: short term introductory workshops, long term project based formats and formal, real-life DT application settings. In Fig. 2, we see effective outcomes (blue) are highest, while both skill-based (green) and cognitive-based outcomes (red) underperform in all of short-term, long term and the real-life settings. Taheri conclusively points out "the threat of neglecting the skill-based outcomes; as this may eventually result in unrealistic expectations about what can be achieved in a DT training and applied afterwards" (Taheri et al. 2014, 2016).

[1]Skill-based outcome: one of the elements of a classification scheme of learning outcomes based on work by Kraiger et al. (1993), Bloom's (1956) and Gagne's (1984), taxonomies which provides guidelines for researchers in training evaluation, taking a multidimensional approach to learning outcomes. Their lens suggests learning as evidenced through the variation in (1) skill-based, (2) affective and (3) cognitive states of trainees.

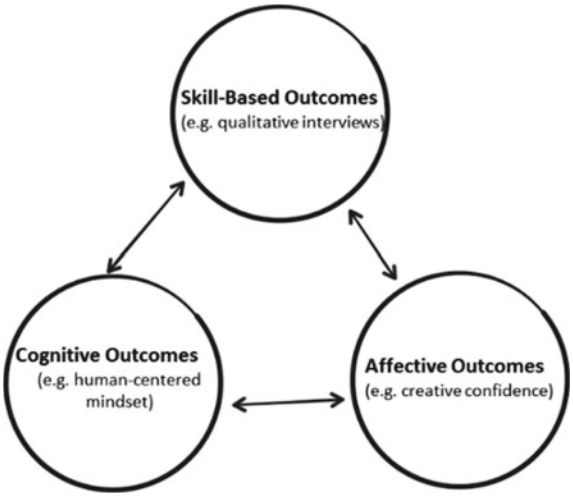


Fig. 1 Conceptual model for design thinking learning domains (Taheri et al. 2016)

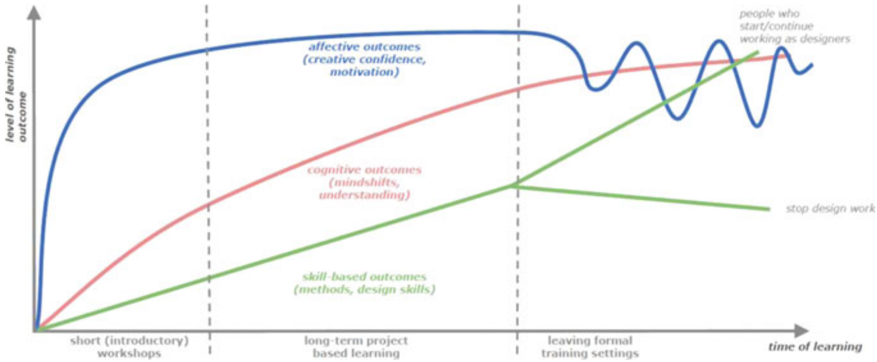


Fig. 2 A conceptual model of expected outcomes of current DT training formats (Plattner et al. 2016; Taheri et al. 2016)

Moreover, Taheri emphasizes "the potential dangers for educational training of future innovators and leaders at the university level." The authors believe that "Neglecting the skill-based outcomes may lead to educating individuals with creative over-confidence, who lack the skills and knowledge to apply their creativity."

In our work as teachers and practitioners, we have found Taheri's insight to be true. To that end, our work takes a *praxis approach* as a remedy for problems in current DT education through emphasizing more skill-based and cognitive-based outcomes.

3 Design-as-Performance: A Praxis Approach to DT Education

> Praxis need not only be seen as the relationship of theory to practice in terms of the work it produces in the studio; praxis can also be used as a means by which students can gain access to what could be described as 'high theory'—(Farrier 2005)

The radical point of view of this research considers the work of design teams to be a performative act (Designing-as-Performance or DaP) moreover, that design sessions are a performance of a corpus of behaviors that constitute much of the practice of Design Thinking.

There are two common perspectives on the word Praxis. The first is the application of the word to mean 'practice.' In this first instance, the Oxford English Dictionary defines Praxis as "formed of habitual action, accepted practice or custom."

The second definition of the word praxis is "an effort of will to transform theoretical concepts and considerations into shared physical activity." Of the two definitions, we are particularly concerned with the second sense of the word praxis. We also employ the extended definition from Farrier, where he describes praxis as a modality where "the relationship between theory and action is played out in the studio setting" (Farrier 2005).

Based on Farrier's model of praxis, we propose Designing as Performance (DaP): a studio setting praxis approach to DT education that not only strengthens both cognitive and skill-based outcomes but also ensures that highly effective performative patterns consist of sound theory and rigorously formulated practice. These cognitive-behavioral models are taught, learned and mastered by individuals as well as teams in redesign scenarios By employing a praxis approach, "the values of theory can be 'embodied' in physical situations" (McCullough 1998).

Taking a praxis approach, therefore, informs our research. Here we align with Farrier who employs the notion of circularity to describe "the relationship of theory to practice that has been developed in several places in the academy and the industry." We agree with Farrier in that we consider the relationship between theory and practice as **circular** (Fig. 3a) and not triangular (Fig. 3b). We also enlist Farrier's cyclical formulation that the relationship is **fluid and dynamic**. Our praxis approach, therefore, leverages on circular formulations without the difficulties that a triangular formulation may have. The quote below from Farrier expresses clearly our notion of the praxis of designing-as-performance.

> Rather than seeing practice and theory at opposite ends of a shape
> that values one over the other; a cyclical relationship can be used to
> describe to what extent theory and practice can be seen as equally interrelated—(Farrier 2005)

In the context of DT education, we see practice as a part of theory and theory as a part of practice; the two do not have mutual exclusivity and are therefore equally important. We do not perceive design theory as descriptive of the practice of design,

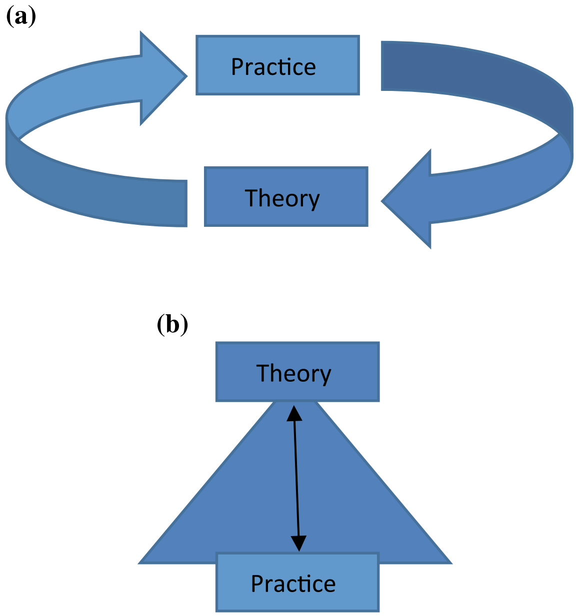


Fig. 3 **a** Circular relationship between theory and practice, a cyclic, fluid and dynamic relationship (Farrier 2005), **b** Triangular relationship between theory and practice, a top down relationship with theory at the pinnacle (Farrier 2005)

but as a dynamic bond to "the creative process which enables a different grade of perception of the possibilities of making performance" (Farrier 2005).

In Designing-as-Performance, we present the creative potential of design theory as a outlining a field of opportunities to DT academics, practitioners and students alike. Complementing current design thinking education with DaP praxis approach involves educating DT learners in understanding DT theory. This would afford DT learners the ability to also critique theory, a process both beneficial for students as well as the design thinking research community. The DaP praxis approach also entails a phenomenological practice of design which takes into account the perception-action loops between the designer's bodies and their environment as well (Edelman and Currano 2011; Kirsh; Kirsh; Rietveld et al. 2018; Edelman 2011). Furthermore, the approach also considers the specific context in which the team is situated as well as the artifacts and media with which the design team interacts (Edelman et al. 2012; Tversky 1993, 2003a, b). Finally, praxis avails both designers and design thinking researchers with an experiential "testing" ground in that the knowledge gained is both intellectual knowledge, as well as physical understanding.

Because an understanding of performance is crucial to our point of view, we provide the following background on performance. We will then frame the structure of the training packages based on training models in performative disciplines like music and sports.

4 Theory of Performance

> The mere act of framing any activity as performance makes it into a performance.
>
> —*John Cage*

It is common knowledge that humans are capable of extraordinary accomplishments, in other words, excellence. These accomplishments are most often produced from a high-level performance (Wilson et al. 2015). Traditionally, performative disciplines have relied on a combination of theory and structured practice that reinforce desirable behaviors which are critical for performative excellence. In the case of sports (Porter 1974; Schmidt and Lee 2014), the understanding of theory and body mechanics as well as the repeated application of this understanding in multiple use scenarios (skills, drills, and free play), are critical for high performance. In the same manner, musical performance (Harnum 2014), enjoys a long tradition of training which is comprised of musical theory, body mechanics, skills, drills, and free play as requirements for outstanding performance. Performance and creativity have also recently been explored in the context of cognition in the theme of dance by Kirsh (Kirsh 2010a, b, 2011a, b; Kirsh et al. 2012).

In our understanding of performance, we build on Erving Goffman's work "The Presentation of Self in Everyday Life" (1959). Goffman defines performing as a behavioral model characterizing any activity. Goffman sees performance as a "quality" that can occur in any situation rather than a fenced-off genre. (Goffman 1959, 1990) We also embrace the composer John Cage's conception of performance. For Cage, "the mere act of framing any activity as performance makes it into a performance" [Schechner and Schechner 1988; Schechner 2003 (2005 printing)].

4.1 Defining Performative Patterns

We note that there are two senses of performance, both of which DaP seeks to cultivate. The first sense is the act of performing itself; the second refers to the results of the performance and points to producing valued results. Performance can take the form of an individual (see Fig. 4a), or a group of people engaging in a collaborative effort (Fig. 4b). There are several examples of performance that we can apply to evaluate what designing-as-performance entails. Generally speaking, music, play, games, sports, theater, and ritual all have "performance" in common.

Before defining performative patterns, we will provide several examples drawn from sports and music. An example of a performative pattern in music occurs in the practice of jazz. Jazz patterns typically use scales, modes, simple chords, complex chords (cycle of fifth, chromatic, stepwise) (Coker et al. ca. 1990). All these jazz patterns help jazz learners improve their hearing ability (listening), develop finger to mind/ear connections, as well as implementing phrase styles. Practicing jazz patterns also helps students to have a deeper understanding of how expert soloists think

(a) (b)

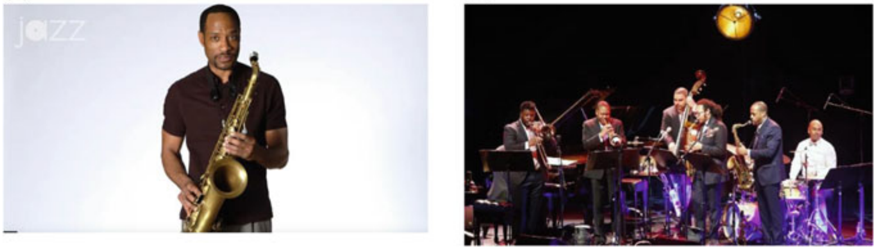


Fig. 4 **a** Walter Blanding emphasizes how practicing scales as an individual is crucial to performance. (jazz.org). **b** Jazz Team with Jazz artist Wynton Marsalis (second from left) performing in concert with his team (Walter Blanding on the Saxophone) (Ravindranath 2017)

through chord progressions, as well as mastering rhythmic constructs used by the best soloists. (Haidet et al. 2017; Cho 2010)

In musical improvisation, a frequently practiced performative pattern is the “call and response,” in which players call out musical phrases to one another and return variations on them. An example of this is “You Rascal You,” performed by Louis Armstrong and Louis Jordan. We note that “call and response” is a performative pattern which acts as a container for improvisation, invention, and execution unfolding in the moment.

Performative patterns in jazz facilitates group communication in co-creating music. They allow the group to stay on the same page and at the same time push the boundaries of the music. In this way, performative patterns in jazz create a shared body of behaviors and knowledge that serve as a container for previously undefined content.

An example of a performative pattern in team sports is the “play” in American Football. Plays are predetermined plans that the team practices repeatedly, they often involve strategic and tactical decisions based on where the ball is situated on the field. Plays often have several alternatives that can be enacted depending on the movements of the opposing team. Thus, the play anticipates a number of un-choreographed possibilities. Thus, like a performative pattern in jazz, the “play” in football is a performative pattern that serves as a container for improvisation, invention, and execution unfolding in the moment.

Figure 5a, b show the Four Verticals Play. If viewers of a match like this are unfamiliar with American Football, they might be inclined to see a group of men shoving one another around, until the Quarterback throws the ball and someone catches it. However, this movement is a choreographed routine that anticipates changes to the routine due to the opposing team’s responses. A performative pattern of this kind allows the team to read one another: it allows the receivers to read the defense within defined boundaries; it allows the quarterback to read the choices the receivers make so he can deliver the ball to a place where no one is at the time of release.

Thus, a working definition of a generic performative pattern is *a set of defined iterative interactions that serve as a container for previously undefined content.*

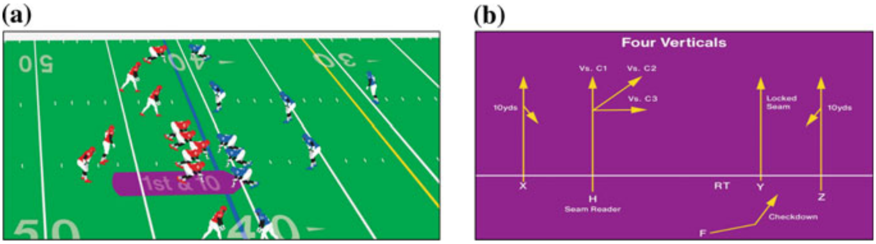


Fig. 5 **a** The four verticals play diagram. **b** The four verticals play on the field

Fluency with performative patterns is one of the factors that allow team members to read one another under quickly changing circumstances because they provide a shared map to contain a range of possibilities for acting on a situation.

In the paper "Teamwork in the Performing Arts" (Rouse and Rouse 2004), Rouse suggests that what makes team performance possible are shared, *mental models*. It is our understanding that Rouse's excellent contribution is based on a Cartesian cognitive model in which thinking happens exclusively in the head. We reframe Rouse's insights in the context of contemporary cognitive models (Kirsh; Tversky 2003a; Clark and Chalmers 1998) and suggest that the models are not merely *mental models*, but *performative patterns*, the elements of which are *theory*, *behaviors* and the *media* which teams use. This is analogous to the patterns, execution, and equipment enlisted in sports (plays; running, passing, catching, scoring; balls, nets, shoes) and music (scales, chords, scores; technique and interacting; instruments).

4.1.1 What Is a Performative Pattern in Team-Based Design?

A performative pattern in design is a set of defined iterative interactions that serve as a container for previously undefined content, that move the inquiry towards potentiality and/or differentiation. Performative patterns in design are often done in groups and mediated with models, tools, and materials.

5 Teaching Performative Patterns

Now that we have established a working definition of performative patterns, we offer a brief survey of some of the methods that sports and music enlist to cultivate expertise in performative patterns.

What follows are examples of several kinds of training from sports and music, though numerous examples can be found in training in any performative discipline. Our proposal is these examples from sports and music provide a model for creating effective curricula for designing as a performative activity.

5.1 Fundamental Units

Practicing scales in music constitute one of the foundations of technical proficiency as well as compositional proficiency. Walter Blanding of the Jazz at Lincoln Center Orchestra (Blanding 2013), relates that he still practices scales after playing for over thirty-five years. Building on basic scales, Blanding suggests that instrumentalists develop their own drills to expand fundamentals. Scales, along with chords and genres, are the building blocks of musical composition (Coker et al. ca. 1990) Thus, the frequent and long term practice of fundamental units like scales can serve as a foundation for performative proficiency, both technical and for development of new material.

In the domain of Classical Music, Chopin's Etudes is an example of highly prescriptive performative instruction. The Etudes enlist specific technical challenges in the context of deep emotional content (meaning). Chopin's insight and contribution is that mastery in music is the joining of the technical and the poetic.

Sports training enlist analogous training practices to those employed in music. An example is the use of *kata* in Judo and other martial arts. A kata is a very formal training method in which the players perform predetermined patterns in order to achieve mastery for application in unstructured matches. Much like the etudes of Chopin, these highly structured exercises are a joining of the technical and the poetic.

Mature disciplines like swimming characteristically deconstruct performance to a remarkably fine granularity. In the following screenshots, we see an account of the physics of buoyancy in breaststroke called "loading" which we see as *swimming theory*; an analysis of the action of the arms in breaststroke; the concept of the "catch" (the critical first part of the stroke), and an exercise called the "front scull" which is a popular exercise practiced to cultivate a proper catch. Furthermore, specific warm-ups and stretches are enlisted that improve flexibility to aid in a range of movement and reduce drag in the water (Fig. 6).

The equivalent elements in music are music theory, understanding genre, chord changes, scales, attack, and phrasing. These are taught for understanding, for an embodiment in playing, and as a ground for communication and development of new musical content.

In music and sports, we have observed theory, repeated practice of foundational units, and repeated contextualization of these foundational units into a broader context of performance to form the core of successful training of high-performance

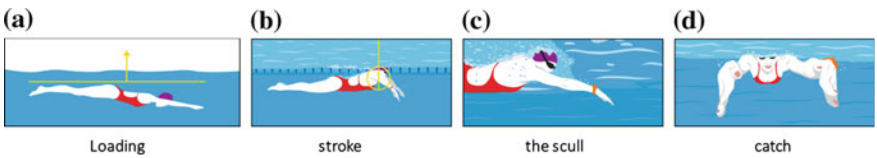


Fig. 6 Swimming: loading, the stroke, the catch and the scull

individuals and teams. In sum, research in musical performance training and sports training suggests:

- Designers may benefit from relevant theory and structured practice of design behaviors
- These behaviors are repeatable and understandable
- These behaviors can be articulated into theory, warm-ups, skills, drills and structured exercises (etudes or kata)
- Repeated practice of well-crafted warm-ups, skills, drills, and exercises build fluency and expertise.

The question we now address is, what characterizes the equivalent to fundamental units in music, plays in team sports, and how can they be taught as skills and etudes or kata?

6 Designing as Performance: Components, Elements, and Core Concepts

A large body of knowledge in Design Thinking and adjacent fields provides the foundational basis for improving design education. A selection of research insights is explored in the following paragraphs, including a summary of the core concepts and their potential impacts on Design Thinking education.

6.1 Media, Behaviors, and Frameworks for Performance

In his doctoral studies, Edelman sought to provide a clear understanding of how the design process works for designers *in situ*, to support their decisions and choices in redesign scenarios (Edelman 2011). Working at Stanford's Center for Design Research (CDR), Edelman's research contextualizes design outcomes in a broader web of behaviors and interactions between teams and the media that they enlist in redesign scenarios. Like other work being done at Stanford's CDR, Edelman sought to identify the empirically observable and measurable characteristics and mechanisms of high-performance teams at work. Acknowledging that "the activity of design is a complex social and technically mediated endeavor" (Jung and Leifer 2011), Edelman's work provides insight into "how design media and behavior entwine to afford the exploration of (sometimes imaginary) worlds and (sometimes imaginary) objects" (Edelman et al. 2012).

His work further presents, an empirically grounded framework to help us understand "how and under what conditions small horizontally organized design teams perform radical redesigns or radical breaks." Edelman focused on designers performing incremental improvements and mid-level redesigns and employed an observation-based case-study approach to examine small design teams in a redesign task. In conclusion, Edelman's work gives evidence for the work of design teams to be an example of extended cognition; design thinking is accomplished through the expert management of concepts, behaviors, and media.

6.2 Accessing Highly Effective Performative Patterns

The title of this chapter is "Accessing Highly Effective Performative Patterns." By "accessing" we mean both identifying and making available often unseen or overlooked design behaviors that, upon examination contribute to robust team performance and meaningful outcomes. In the first sense of accessing, *qua* identifying, research itself has provided a substantial and growing body of knowledge that identifies the fundamental elements of effective team interactions, whether they be the kinds of questions designers ask (Eris 2003), the kind of new language designers create (Mabogunje 1997), the kind of media they enlist (Edelman 2011), or the kind of frameworks they use to structure and move through a redesign activity (Edelman 2011).

The second sense of accessing, making research insights available, is the work of bringing research to impact. This entails the creation and validation of training packages that translate new knowledge into actionable materials that designers can use. Based on our survey of training methods in performative disciplines, we have formulated eight formal elements which constitute a training package.

6.3 Formal Elements of Training Packages

1. Theory
2. Warm-Ups (curated, simple activities to gain familiarity with the performance of concepts)
3. Individual Skills (in musical terms "chops,")
4. Team Drills (these develop a clear sense of team roles and interactions)
5. Scripted Practice (following the approach of Chopin and the Martial Arts, structured Etudes or Kata; these can be done repeatedly to fine tune skills)
6. Speed Drills (timed interval training, to cultivate quick responses)
7. Free Play (exercises to build fluency, for design teams to experiment with the performative pattern)
8. Toy (an advanced, full redesign exercise used for assessment).

In practice, we have introduced each training package with Warm-Ups, rather than beginning with Theory. The thinking behind this is to provide a carefully curated short experience that gives the essence of the performative pattern *qua* performance. This is followed by a pattern of instruction which combines theory and exercises in increasing detail. In this way, we strive to co-develop skill and knowledge at the same time.

Our research group has several research-based training packages in development. Many of these have been tested in several venues both in Europe and the United States. The packages are based on the following research:

- Media Models: the media that designers enlist have cognitive affordances (Edelman 2011)
- Generative Design Questions and Deep Reasoning Questions: two kinds of questions that designers use to frame effective inquiry (Eris 2003)
- Solicitations: designers create models that solicit phase appropriate actions (Rietveld et al. 2018)
- Noun Phrases: designers create new language (Mabogunje 1997)
- Dimensions of Engagement: a systems approach to generative product service architecture (Edelman 2011)
- Disruption-Integration: the master algorithm (Edelman 2011; Menning et al. 2018)
- Enactment: acting out in semi-imaginary worlds (Edelman and Currano 2011)
- Marking: designers enlist a shorthand sketch for enacting interaction (Kirsh 2011b)
- Remapping: transposing touch points on to different form factors for new usability and use-cases (Edelman et al. 2012; Edelman 2011)
- Four Forces of Change: Aristotle's Four Causes in the service of design (Edelman and Currano 2011; Edelman 2011)
- Metaphor: using metaphor to leverage high impact opportunities (Edelman and Currano 2011; Edelman 2011).

To include an account of each and every training package mentioned above is beyond the scope of this chapter. However, we offer three parts of a training package based on Ozgur Eris' work on the kinds of questions design teams ask as an example of how an actual training package looks. The entire training package has these elements:

(1) *Theory:* high-performance design teams ask two kinds of questions: GDQs and DRQs
(2) *Warm-Ups*: about X ask questions, ask DRQs, ask GDQs, alternate, answer
(3) *Individual Skill*: ask GDQs and DRQs with X, ask and answer GDQs and DRQs with X
(4) *Team Drill*: ask GDQs and DRQs with X in turn, ask and answer GDQs and DRQs in turn
(5) *Scripted Practice*: design team practices a design Etude or *kata* of GDQs and DRQs
(6) *Speed Drills*: design team practices asking and answering GDQs and DRQs in turn in 10 and then 5-s intervals
(7) *Free Play*: design team ask and answer GDQs and DRQs freely
(8) *Toy*: design teams work to expand horizons on realizing an advanced, complete redesign

For the sake of demonstration, we now present three parts of a package based on Eris' Generative Design Questions and Deep Reasoning Questions (DRQs & GDQs). We will first describe a Warm-Up, then Theory, and finally a Speed Drill.

GDQs & DRQs Warm-Up: Asking Questions

The purpose of these Warm-Ups is to train designers to be sensitive to the questions they ask. This series of Warm-Ups begin as (1) general questions, then (2) questions that move a narrative forward, then (3) cultivate to specific kinds of questions, and finally, (4) specific questions with appropriate answers.

Team members gather in a circle and

(1) Ask questions in turn without responding with answers, just questions
(2) Ask questions that move the narrative forward (e.g., "What are you eating?" "Would you like some?" "Aren't you hungry?")
(3) Ask specific kind of questions that concern an object or experience (e.g., planning an event like a party)
    (a) specification, comparison, and verification (e.g., just how many people are we inviting?)
    (b) what ifs (e.g., what if we all wore togas?)
(4) Same as above but with appropriate answers.

GDQs and DRQs Theory: Asking the Right Questions at the Right Time

Ozgur Eris studied the kinds of questions that designers ask when they are working in teams. Eris found that a combination of Deep Reasoning Questions and Generative Design Questions are needed for successful design outcomes (Eris 2003).

Deep Reasoning Questions (DRQs) are concerned with verification, comparison, specification, in other words, logical status:

- Verification (Is this true?)
- Comparison (Is this heavier or lighter?)
- Specification (Just how big is this?)

Generative Design Questions (GDQs) are not concerned with verification, comparison, specification. Instead, they are concerned with generating possibilities:

- Proposal/Negotiation (How about attaching a wheel to the long LEGO piece?—aimed at establishing a negotiation process based on opinions)
- Scenario Creation (What if the device was used on a child?—aimed at generating a multitude of outcomes)
- Ideation (Are magnets useful in any way?—aimed at generating a multitude of concepts)
- Method Generation (How can we keep the wheel from slipping?—aimed at generating secondary conceptualizations)
- Enablement (What allows you to measure distance?—aimed at identifying resources).

Eris' work drills deep into exactly how questions frame the outcomes of inquiry. In practice, these excellent and fundamental distinctions concerning Generative Design Questions are challenging to master in the short term. We have experimented with GDQs in our workshops and by way of introduction have essentialized GDQs into these three questions:

- How might we?
- What are the ways we could?
- What kinds of scenarios could we imagine?

Additionally, Eris' work highlights a common cause of team dysfunction. We have often witnessed teams that are unaware of what kinds of questions they are asking and unaware of the impact that the questions have on the direction of the team. We have also witnessed occasions in which team members are unwittingly asking DRQs and GDQs at the same time and as a result, are growing frustrated because they are at loggerheads and going nowhere.

GDQs & DRQs Speed Drills: Timed Interval Training

Team members ask DRQs and GDQs in several rounds. Each round is timed for each object and team member in turn; round one is six seconds per turn, round two is four seconds per turn, round three is two seconds per turn.

(1) *Ask DRQs (specification, comparison, verification) about an object X*
(2) *Ask GDQs (how might we?, what are the ways we could?, what kinds of users or scenarios?) about an object X*
(3) *Alternate 2 and 3 above*
(4) *Same as 2, 3 and 4 above with answers*

Where X is a bottle, a camera, planning a birthday or workshop.

6.4 Assessing Training Packages

We generally get good feedback from designers we have trained. What follows is a selection of comments from students and professionals that have participated in our workshops.

> The research-based exercises allow me to understand exactly where the re-design challenge is located in the process… and where I can start a disruption or change
>
> The exercises…were so understandable, exercising felt very intuitive and logical.
>
> It felt like having an x-ray of creative working sessions.
>
> As a Design Thinking Coach and professional designer it was a very helpful workshop to understand how to teach Design Thinking with more concrete and precise exercises.
>
> The methods can be used to find out who is the best "point guard", the best "center" and so on. But they can also be used to make "centers" into "point guards" or the other way round - in a very structured, thus protected framework…This way, people can try out roles that they or others hadn't foreseen for themselves.

Table 1 Performative pattern assessment protocol

Pre-training assessment	Training	Post-training assessment
KAI Creativity test	Single exercises	KAI Creativity test Video recorded redesign task with physiological data
Video recorded redesign task with physiological data Self reporting for creative confidence	Single packages	
Semi-structured interview Expert evaluation of outputs	Multiple packages	Self reporting for creative confidence Semi-structured interview Expert evaluation of outputs

> The workshop was really a massively pivotal point for me in my way of thinking about design.

As gratifying as positive feedback can be, more objective assessments are necessary to ensure rigor. While some of our research into assessment is truly a work in progress, there are several assessments that we are starting to implement. Table 1 is a schematic of the assessment protocol we are currently implementing.

We plan to assess multiple teams of three participants. A control group will be trained with standard Design Thinking materials. Another group will be trained with the research-based training materials that we have described above. Our interest is in determining which approach is more effective in cultivating creativity, high-quality outputs from the redesign task, qualitative assessment regarding participants' sense of their creative confidence, and gathering data concerning the physiological state of participants engaged in a team-based redesign task. The last of these, gathering physiological data is a new enterprise, in part directed toward seeking insights and objective information about how people feel when they are designing, and in part meant to be a complement to new work being done in NeuroDesign at Stanford.

7 Conclusion

We have observed that much of Design Thinking instruction is ten years behind in embodying and communicating new knowledge about design and design teams. Research has matured the discipline of Design Thinking beyond a loosely connected set of best practices. However, the new knowledge resulting from over ten years of research has seldom been implemented in the form of teaching and learning materials.

This paper has presented a new approach to designing and design education called Designing as Performance, which involves not only the intellectual task of designing and understanding theory but a phenomenological practice of perception-action loops between the body, the environment in which the team is situated and the artifacts with which the team interacts. The work of design teams is a performative act, and that design sessions are a performance of a corpus of behaviors that constitute much of the practice of Design Thinking. This corpus of behaviors is repeatable and understandable and thus can be trained and learned in the form of a skills repertoire, the core of which are performative patterns.

Seeing the world as a field of possibilities and acting on the world with skilled action is the work of design. We hope that the research and training in which we are engaged has enabled designers to see the world as a field of possibilities, and has moved and will continue to move designers to act on the world with thoughtful, reflective and skilled action.

Publication Bibliography

Blanding, W. (2013). *How to practice scales in Jazz—YouTube*. Available online at https://www.youtube.com/watch?v=Anq26tezBSk&t=26s. Checked on February 18 2, 2019.

Bloom, B. S. (1956). *Taxonomy of educational objectives*. Handbook: The Cognitive Domain. New York: David McKay.

Bloom, B. S. (Ed.). (1987). Taxonomy of educational objectives. The classification of educat. goals; handbook. 30th print. London: Longman Group.

Cho, A. (2010). *The jazz process. Collaboration, innovation, and agility*. Upper Saddle River, NJ: Addison-Wesley.

Clark, A., & Chalmers, D. (1998). The extended mind. In *Analysis* (Vol. 58, no. 1, pp. 7–19). Available online at http://www.jstor.org/stable/3328150.

Coker, J., Casale, J., Campbell, G., & Greene, J. (ca. 1990). *Patterns for Jazz*. For treble clef instruments. 3. ed. Miami, Fla.: Studio P/R.

Edelman, J. (2011). *Understanding radical breaks: Media and behavior in small teams engaged in redesign scenarios*. Available online at https://purl.stanford.edu/ps394dy6131. Updated on January 29, 2019, checked on February 18, 2019.

Edelman, J. (2019). *Accessing highly effective performative patterns in team-based design*. Dartmouth College, 2019. Available online at https://engineering.dartmouth.edu/events/accessing-highly-effective-performative-patterns-in-team-based-design. Checked on February 18 2019.

Edelman, J., Agarwal, A., Paterson, C., Mark, S., & Leifer, L. (2012). Understanding radical breaks. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research. Studying co-creation in practice* (pp. 31–51). Heidelberg: Springer (Understanding innovation).

Edelman, J., & Currano, R. (2011). Re-representation: Affordances of shared models in team-based design. In H. Plattner, C. Meinel, & L. J. Leifer (Eds.), *Design thinking. Understand—improve—apply* (pp. 61–79). Heidelberg, London: Springer. Available online at https://doi.org/10.1007/978-3-642-13757-0_4.

Eris, O. (2003). *Asking generative design questions: A fundamental cognitive mechanism in design thinking*.

Farrier, S. (2005). Approaching performance though praxis AU. *Studies in Theatre and Performance, 25*(2), 129–144. https://doi.org/10.1386/stap.25.2.129/1.

Gagné, R. M. (1984). Learning outcomes and their effects: Useful categories of human performance. *American Psychologist, 39*(4), 377–385. https://doi.org/10.1037/0003-066X.39.4.377.

Goffman, E. (1959). *The presentation of self in everyday life*. 1st Anchor books edition. New York: Anchor Books.

Goffman, E. (1990). *The presentation of self in everyday life* [Nachdr.]. New York, NY: Doubleday (Anchor books).

Haidet, P., Jarecke, J., Yang, C., Teal, C., Street, R., L., & Stuckey, H. (2017). Using jazz as a metaphor to teach improvisational communication skills. In *Healthcare (Basel, Switzerland)* (Vol. 5, no. 3). https://doi.org/10.3390/healthcare5030041.

Harnum, J. (2014). *The practice of practice*. [Online content edition]. [Chicago, Illinois]: Sol UT Press.

Johansson-Sköldberg, U., Woodilla, J., & Çetinkaya, M. (2013). Design thinking: past, present and possible futures. *Creativity and Innovation Management, 22*(2), 121–146. https://doi.org/10.1111/caim.12023.

Jung, M., Leifer, L. (2011). A method to study affective dynamics and performance in engineering design teams. In *ICED 11–18th International Conference on Engineering Design—Impacting Society Through Engineering Design* (Vol. 7).

Kelley, D. (2018). *The little book of IDEO*. Available online at https://lboi.ideo.com/. Updated on May 1, 2018, checked on February 14, 2019.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, *78*, 311–328. https://doi.org/10.1037/0021-9010.78.2.311.

Kirsh, D. (1996). Adapting the environment instead of oneself. *Adaptive Behavior*, *4*(3–4), 415–452. https://doi.org/10.1177/105971239600400307.

Kirsh, D. (2000). A few thoughts on cognitive overload. *Intellectica. Revue de l'Association pour la Recherche Cognitive*, *30*(1), 19–51. https://doi.org/10.3406/intel.2000.1592.

Kirsh, D. (2006). Implicit and explicit representation. In *Encyclopedia of cognitive science*. https://doi.org/10.1002/0470018860.s00166.

Kirsh, D. (2010a). Thinking with external representations. *AI & Society, 25*(4), 441–454. https://doi.org/10.1007/s00146-010-0272-8.

Kirsh, D. (2010b). *Thinking with the body*. Available online at https://philpapers.org/archive/KIRTWT.pdf.

Kirsh, D. (2011a). *Creative cognition in choreography.* Available online at https://philpapers.org/archive/DAVCCI.pdf.

Kirsh, D. (2011b). *How marking in dance constitutes thinking with the body*. Available online at https://philpapers.org/archive/KIRHMI.pdf.

Kirsh, D., Caballero, R., & Cuykendall, S. (2012). *When doing the wrong thing is right*. Available online at https://philpapers.org/archive/KIRWDT.pdf.

Kolko, Jon. (2010). Abductive thinking and sensemaking: The drivers of design synthesis. *Design Issues, 26*(1), 15–28. https://doi.org/10.1162/desi.2010.26.1.15.

Kolko, J. (2015). *Design thinking comes of age*. https://www.facebook.com/HBR. Available online at https://hbr.org/2015/09/design-thinking-comes-of-age. Checked on February 14, 2019.

Mabogunje, A. (1997). Noun phrases as surrogates for measuring early phases of the mechanical design process. In *Proceedings of the 9th International Conference on Design Theory and Methodology, 1997*.

McCullough, C. (1998). *Theatre praxis: Teaching drama through practice*. Edited by null (null, null).

Menning, A., Ewald, B., Nicolai, C., & Weinberg, U. (2018). *"… and not building on that": The relation of low coherence and creativity in design conversations* (pp. 195–213).

Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2012a). *Design thinking research*. Berlin, London: Springer (Understanding innovation).

Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2018). *Design thinking research. Making distinctions: collaboration versus cooperation*. Cham, Switzerland: Springer (Understanding innovation).

Plattner, H., Meinel, C., & Leifer, L. J. (2011a). *Design thinking. Understand—improve—apply*. Heidelberg, London: Springer.

Plattner, H., Meinel, C., & Leifer, L. J. (Eds.). (2011b). *Design thinking. Understand—improve—apply*. Heidelberg, London: Springer.

Plattner, H., Meinel, C., Leifer, L. J. (2012b). *Design thinking research. Measuring performance in context*. Heidelberg, New York: Springer (Understanding innovation).

Plattner, H., Meinel, C., & Leifer, L. J. (Eds.). (2016). *Design thinking research. Taking breakthrough innovation home*. Cham: Springer (Understanding Innovation, 2197–5752).

Porter, P. (1974). *Judo from the beginning* (Vol. 1). National Coaching Standards, Zenbei.

Ravindranath, K. (2017). Review: Jazz giant Wynton Marsalis edged out of his comfort zone by empathetic duet with oud player Naseer Shamma at Abu Dhabi Festival. Jazz artist Wynton Marsalis (second from left) performing in concert with his team at Emirates Palace as part of the Abu Dhabi festival in Abu Dhabi. The National. Available online at https://www.thenational.ae/arts-culture/review-jazz-giant-wynton-marsalis-edged-out-of-his-comfort-zone-by-empathetic-duet-with-oud-player-naseer-shamma-at-abu-dhabi-festival-1.48940?videoId=5766484581001. Checked on February 14, 2019.

Rietveld, et al. (2018). *Ecological-enactive cognition as engaging with a field of relevant affordances: The skilled intentionality framework (SIF)*.

Rouse, W., Rouse, R. (2004). Teamwork in the Performing Arts. In *Proceedings of the IEEE* 92. https://doi.org/10.1109/jproc.2004.825880.

Schechner, R. (2003 (2005 printing)). *Performance theory*. Rev. and expanded ed. London: Routledge (Routledge classics).

Schechner, R., & Schechner, R. (1988). Essays on performance theory. In *Performance theory*. Rev. and expanded ed. London: Routledge.

Schmidt, R. A., & Lee, T. D. (2014). *Motor learning and performance. From principles to application* (5th ed.). Champaign IL: Human Kinetics.

Sonalkar, N., Mabogunje, A., & Cutkosky, M. (2018). Quadratic model of reciprocal causation for monitoring, improving, and reflecting on design team performance. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research. Making distinctions: collaboration versus cooperation* (Vol. 14, pp. 43–57). Cham, Switzerland: Springer (Understanding innovation).

Taheri, M., Unterholzer, T., Hölzle, K., & Meinel, C. (2016). An educational perspective on design thinking learning outcomes. Available online at https://hpi.de/fileadmin/user_upload/fachgebiete/meinel/papers/Design_Thinking/2016_taheri_unterholzer_ispim.pdf. Checked on June 1, 2019.

Taheri, M., Unterholzer, T., & Meinel, C. (2014). *Design thinking at scale. Investigating the MOOC potential of design thinking training*. Available online at https://hpi.de/en/meinel/knowledge-tech/innovation-research/design-thinking-at-scale.html. Updated on January 1, 2018, checked on February 18, 2019.

Tversky, B. (1993). Cognitive maps, cognitive collages, and spatial mental models. In A. U. Frank, I. Campari (Eds.), *Spatial Information Theory. A Theoretical Basis for GIS: European Conference, COSIT '93,* Marciana Marina, Elba Island, Italy September 19–22, 1993: Proceedings (Vol. 716, pp. 14–24), Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg.

Tversky, B. (2003a). Navigating by mind and by body. In C. Freksa et al. (Eds.), *Spatial cognition III. Routes and navigation, human memory and learning, spatial representation and spatial learning* (Vol. 2685, pp. 1–10), Lecture notes in computer science, Lecture notes in artificial intelligence 0302-9743, 2685. Berlin, London: Springer.

Tversky, B. (2003b). Structures of mental spaces. *Environment and Behavior, 35*(1), 66–80. https://doi.org/10.1177/0013916502238865.

von Thienen, J. P. A., Clancey, W. J., Corazza, G. E., & Meinel, C. (2018). Theoretical foundations of design thinking. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research. Making distinctions: Collaboration versus cooperation* (Vol. 8, pp. 13–40). Cham, Switzerland: Springer (Understanding innovation).

Wilson, M., Sleutel, M., Newcomb, P., Behan, D., Walsh, J., Wells, J. N., et al. (2015). Empowering nurses with evidence-based practice environments: Surveying Magnet®, Pathway to Excellence®, and non-magnet facilities in one healthcare system. *Worldviews on Evidence-Based Nursing, 12*(1), 12–21. https://doi.org/10.1111/wvn.12077.

Designing a Synthesis MOOC: Lessons from Frameworks, Experiments and Learner Paths

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

Abstract We are constantly striving towards improving our Massive Open Online Courses (MOOCs), in which we convey design thinking skills to thousands of learners. In this chapter we describe how lessons learned from our first MOOC and different educational frameworks influenced the course design for our second MOOC on synthesis, idea generation and idea selection. We present general survey results from the second course as well as a preliminary analysis and discussion of the synthesis assignment task.

1 Introduction

Teaching design thinking skills in Massive Open Online Courses (MOOCs) enables us to introduce learners worldwide to new methods and mindsets. The challenge to conceptualize and run these online courses also requires us as (design thinking) educators to experiment with instructional strategies beyond the traditional workshop format. In our course design, we take an approach that is modular, task-based, and inspires skill application. The modularity of our course system allows learners to explore any of the three course topics separately. The four to six week courses focus on building empathy, synthesis and idea generation, and prototyping and testing respectively. The aim of the modular approach is to encourage learners who want to refresh certain skills or who are particularly interested in a specific phase. In October 2018, we ran the second part of our MOOC series on design thinking skills:

L. Mayer (✉) · K. von Schmieden · M. Taheri · C. Meinel
Hasso Plattner Institute for Digital Engineering, Prof.-Dr.-Helmert Straße 2-3, 14482 Potsdam, Germany
e-mail: Lena.Mayer@hpi.de

K. von Schmieden
e-mail: Karen.Schmieden@hpi.de

M. Taheri
e-mail: Mana.Taheri@hpi.de

C. Meinel
e-mail: meinel@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_3

Table 1 General structure and material of the "Human-Centered Design: From Synthesis to Creative Ideas" MOOC

Weeks	Topic	Material		
Week 1	Introduction	Videos and texts	Introduction game	
Week 2	Synthesis	Videos and texts	Exercise	Peer-reviewed assignment
Week 3	Idea generation	Videos and texts	Exercise	
Week 4	Wrap up	Videos, texts, material for download		

"Human-Centered Design: From Synthesis to Creative Ideas." The course objective was to introduce learners to helpful skills in advancing from user research to idea generation. During the course, learners interpreted research findings, framed user-centered problem statements, and facilitated their own brainstorming sessions. The learning objective involved:

- sorting data and extracting needs from user research
- framing a problem from a user's perspective
- applying various brainstorming techniques and principles
- selecting and conceptualizing ideas
- evaluating the quality of your peers' work and providing constructive feedback.

Table 1 shows the structure and material of this MOOC.

Comparison and changes from MOOC #1 to MOOC #2.[1]

Learnings from our first design thinking MOOC led us to add and adjust aspects both on the content and structure level of our second MOOC.

At the beginning of the course, we presented a short *video on course ethics and values*, with the aim of establishing an environment of mutual respect among learners (Ginsberg 2005). This video covered various topics, including mindfulness of diversity within the course and emphasis on valuing peer feedback over grades. In this way, we addressed the challenges from the first MOOC when learners reported that the level of constructive feedback in peer reviews could be improved. Moreover, different cultures show different attitudes towards feedback (Bailey et al. 1997).

Second, we provided learners with a *question flowchart*. Our aim was to guide learners more efficiently when they faced a problem. We called attention to different resources or contact points for technical issues or content problems, and furthermore updated our "bug list" for possible platform problems.

To better integrate learners in the course, we turned our *literature recommendation list* into an "open source" resource. Course participants were thereby invited to add their own recommendations. In this way, we also hoped to call upon the learners' prior knowledge and to spark a discussion on the quality of different resources.

[1]MOOC #1 (start: September 2017) was called "Inspirations for Design: A Course on Human-Centered Research" and we taught observation and qualitative interviewing skills.

MOOC #2 (start: September 2018) was called "Human-Centered Design: From Synthesis to Creative Ideas" and targeted synthesis, idea generation and idea selection.

We experimented with a structure adjustment, allowing *more time flexibility* for learners. The content for synthesis and idea generation as well as selection was accessible simultaneously, and we incorporated both topics in one assignment. This helped to reduce the number of deadlines, provide more time for the assignment and reduce peer reviewing tasks.

Finally, we aimed at providing more of a theoretical foundation for learners who wanted to explore the scientific context of design thinking skills. As we also always try to offer content through different media and to target different learning preferences (such as visual, auditory and kinesthetic), we created two *podcasts* that delved into the theory of design synthesis and brainstorming.

To explain such additions, in this chapter we discuss learning design frameworks that inspired our course design. We also present results from the course and exemplify design thinking novice learner behavior with a random sample from the synthesis assignment.

2 Course Design: Inspired by Frameworks and Experimentation

In our course design, we build on established educational frameworks and experiment with new interventions and approaches. In the following, we discuss several learning design frameworks and how they impacted the design of our second MOOC.

A. *Conole's "Seven C" framework* (2014)
B. *Merrill's "Pebble in the Pond" framework* (2002)
C. *Merrill's First Principle of Instruction* (2002)
D. *Bloom's Taxonomy* (1994)
E. *Chickering and Gamson* (1987)
F. *Universal Design For Learning* (*UDL*)
G. *Culturally Responsive Teaching* (*CRT*)

A.–C. *Conole's "Seven C", Merril's Pebble in the Pond, and Merril's First Principle of Instruction*

There are a number of frameworks that deal with the strategic design of learning units. *Conole's "Seven C" framework* (2014) illustrates the key stages in course design, from conceptualization to evaluation in the learning context. *Merrill's "Pebble in the Pond" framework* (2002) encourages instructors to identify the initial problem they want learners to solve, and to evolve their course design from there. It is based on *Merrill's First Principle of Instruction* (2002), a set of principles for course design that increases student learning when applied. These principles are "problem-centered, activation, demonstration, application, and integration." The framework was adopted for the MOOC context and expanded by Margaryan et al. (2015), who added the principles of "collective knowledge, collaboration, differentiation, authentic resources, and feedback."

We considered these principles by connecting skills to real-world design thinking examples ("problem-centered"), motivating students to recall and share prior knowledge ("activation"), including demonstration videos of new skills ("demonstration"), asking students to conduct design tasks in their own context for assignments ("application"), and encouraging discussion on transferring the newly learned skills into their working or private life in the forum ("integration").

D. *Bloom's Taxonomy*

Bloom's Taxonomy (1994) structures learning objectives into a hierarchy. This allows educators to categorize and phrase learning objectives for their own courses (Krathwohl 2002). The original taxonomy represented a cumulative pyramid hierarchy of six categories in the cognitive process domain, ranging from "remember, understand, apply, analyze" to "evaluate" and "create"—create being the highest level. The revised version of the taxonomy separates the knowledge and the cognitive process spectrum. The horizontal axis contains the six hierarchical thinking skills, while the vertical axis shows the four categories of the knowledge dimension. In this way, objectives can be represented in a two-dimensional taxonomy table (Krathwohl 2002).

In our course, we aimed towards learning objectives in the higher order thinking level by designing exercises and assignments that encouraged learners to apply their learning to their own context and report on the results.

E. *Chickering and Gamson's Seven Principles for Undergraduate Education*

The seven principles by Chickering and Gamson (1987) guide educators to assure high quality teaching (Bali 2014). According to the authors, there are seven principles that are essential for good practice in undergraduate education:

1. Encourage contact between students and faculty
2. Develop reciprocity and cooperation among students
3. Encourage active learning
4. Give prompt feedback
5. Emphasize time on task
6. Communicate high expectations
7. Respect diverse talents and ways of learning

Apart from their application in traditional course design, these principles can guide the instructional design of online courses as well (Siemens and Tittenberger 2009). The principles are reflected in different aspects of our MOOC. We monitored the discussion forums closely and promptly answered questions that were addressed to us ("encourage contact between students and faculty"). In some cases, we left space for other learners to respond and thus encouraged cooperation among students. Moreover we used qualitative peer-reviewed assignments, with emphasis on constructive feedback, to promote cooperation among the learning community. Active learning was enabled through exercises and assignments that focused on the application of skills. Along with a visual timeline, we communicated the estimated time for each week and the activities it contained. We tried to communicate high expectation through the use of task-based assignments and exercises, instead of multiple-choice tests. Moreover, participants would receive a record of achievement only after gaining more than 50% of the graded assignments. We enforced the last principle ("respect diverse talents and ways of learning") by offering different learning materials: videos, podcasts, written summaries and visual materials. We furthermore offered a wide range of choices for learners in the assignment.

The principle of providing prompt feedback remains an ongoing challenge in the context of MOOCs, where thousands of participants require feedback simultaneously. We try to react to questions and problems as quickly as possible. Moreover, when students report a review they received by a peer, the teaching team serves as a second scorer and adjusts the review if necessary.

F. *Universal Design for Learning (UDL)*

Universal Design for Learning (UDL) is a framework that helps educators to design instructional strategies for a variety of learner needs. The framework encourages instructors to offer multiple means of representation (the "what" of learning), action and expression (the "how" of learning), and engagement (the "why" of learning), (Chita-Tegmark et al. 2011).

We incorporated these three principles in the instructional design of our MOOC. The principle of a multiple means of representation was enforced by providing various ways of sharing content with the learners, such as videos, text, and audio. Regarding multiple means of action and expression, we encourage learners to complement the assignment templates with drawings or pictures from their own context, or to modify the templates according to their needs. Finally, regarding the last principle, the introduction game in the first week is an example of how we try to arouse learners' interest and keep participants engaged. Moreover, we provide options for learners to pick and execute their assignments.

G. *Culturally Responsive Teaching (CRT)*

One of the final aspects we addressed in the MOOC design is learner diversity in terms of cultural backgrounds. *Culturally Responsive Teaching (CRT)* refers to pedagogical practices that take into account the cultural diversity among learners. CRT encourages instructors to view learners' diversity as a strength rather than a barrier to learning (Kieran and Anderson 2018). Since different aspects of culture have an impact on teaching and learning (Gay 2002), it is crucial for MOOC designers to consider the great diversity in their learning community and design courses that resonates with learners from different cultural backgrounds.

We paid careful attention to cultural inclusivity while designing and running the MOOC. We applied several practices. For instance, we made sure to use examples and cases from different parts of the world, we used simple language, and refrained from using jargons and domain-specific language, and we shot a video on MOOC ethics, in which we prompt our learners to give constructive feedback to their peers and to be mindful of their language when giving feedback.

In the post-course survey we asked learners to rate the MOOC's performance with regard to cultural inclusivity, and asked if they perceived any aspect of the MOOC as insensitive towards their culture. The results showed that the course performed well in terms of cultural inclusivity.

3 MOOC #2: General Results

"Human-Centered Design: From Synthesis to Creative Ideas" ran from September to October 2018. 3641 learners enrolled during the course, of which 1945 (53%) were active learners who visited the course at least once. 1202 learners used the discussion forum and 526 learners posted in the forum.

According to platform data, most enrolled learners participated from Germany (24.01%), followed by the United States (1.48%) and Switzerland (1.02%). Overall, learners from 69 countries took part in the MOOC. With 53% of survey participants identifying as "male", with slightly fewer female participants attending the course.

77.57% of pre course survey participants ranked their prior experience with design thinking as "none" or "beginner" (163 and 570 respectively out of n = 945). 72% of survey participants had previous experiences with MOOCs, and 24% had even participated in more than 5 MOOCs. Only 262 out of n = 946 were first time MOOC participants.

Overall, survey participants responded positively to the course quality (see Fig. 1).

Out of n = 284 participants in the post-course survey, 256 (90.14%) ranked their satisfaction with the overall course with 7 or above on a 10-point Likert response scale.

257 survey participants responded with "yes" to "Were your learning expectations met?", which amounts to 90.49% of the survey sample.

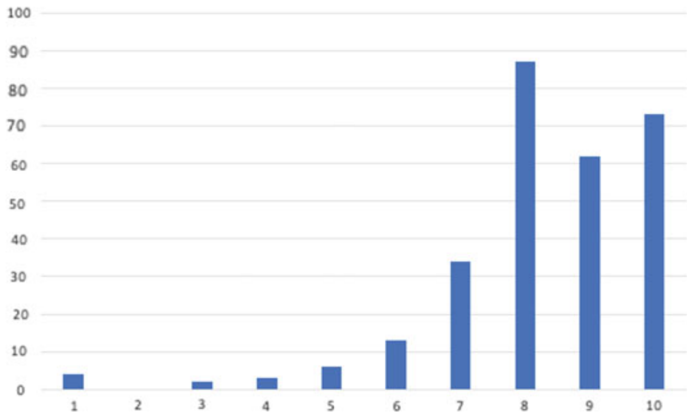


Fig. 1 Participants rated their satisfaction with the overall course. The *x*-axis illustrates the satisfaction answer categories from 1 (not satisfied at all) to 10 (absolutely satisfied). The *y*-axis represents the percentage of participants per answer category

After the official course end, MOOC #2 is available in archive mode,[2] and learners continue to enroll and examine material. We are, however, no longer collecting survey results.

4 Looking into Synthesis Paths

Out of the 1945 active learners, 392 learners successfully participated in the course assignment. We offered the synthesis and ideation MOOC as a modular entity, which allowed participants to learn synthesis skills without doing their own user research first. We therefore provided participants with user research material, or "Case Material" in the MOOC.

For the assignment, all learners received the same user research material in undertaking the task to reframe the user problem in a synthesis process. With their new problem statement, they conducted a brainstorming session and summarized the idea their team chose.

The constrained amount of user research material and the identical starting condition for all participants offers us the possibility to look into the different synthesis paths that learners chose. We can observe a synthesis process at scale and draw learning for the on-site teaching of synthesis.

Available Case Material

To offer a realistic case situation, we presented user research material from a student project at the HPI School of Design Thinking. We anonymized or changed the information in the data to protect the participant's identity. Some quotes had to also be omitted for confidentiality reasons. Nonetheless, the interview quotes and user research notes stayed true to the original data. They represented the main findings of the design team and allowed participants to examine different problems, pain points, and tensions.

The design challenge was to "redesign the security check experience at airports." During the security check, passengers and their hand luggage are examined for potentially dangerous items, such as weapons or liquids. Passengers need to put their hand luggage into boxes which are scanned, as they move through a metal detector and must usually pass through a body scanner. Airport security officers are in charge of sorting the boxes, examining the luggage, conducting the metal detection or body scans, and interacting with passengers regarding their luggage and potentially banned objects. The data was gathered at German airports. The process might differ slightly in airports in other parts of the world.

The case material consisted of explorative qualitative interviews, passive and participatory observation notes, immersion notes, and the result of an empathy card deck sorting by a user.

[2] https://open.hpi.de/courses/ideas2018.

For the qualitative interviews, participants received quotes from four interviewees. Lara, a businesswoman and frequent traveler, was unhappy with the slow process and the incompetence of fellow travelers. Juan, an infrequent traveler, was stressed out by the pace of the security check and felt intimidated by the process (see Fig. 2). Michael, an airport security officer, voiced communication problems with travelers, and did not feel valued by other airport staff in his position. Alima, another airport security officer, viewed her job as stressful and did not feel like passengers took her seriously.

An empathy card deck sorted by frequent traveler and family father Lars showed that he did not associate airport security officer uniforms with "security." A user journey depiction (see Fig. 3), observation and immersion notes, and observation pictures described the security check procedure.

Task

Participants picked one of the users in the case material, defined their "golden nugget," interpreted the user's hidden need or motivation behind the quote or note they chose, and phrased a "How Might We Question" or new problem statement from the user perspective based on their inference. To enable the review by peers, participants also filled in two fields explaining why they picked this specific "golden nugget" and why they phrased the "How Might We Question" in the way they did (see example template in Fig. 4).

Sample

We did a random sampling of 22 assignment samples taken from all participants who scored points (total n = 392). The sample represents learners who uploaded their assignment early on, midway, or shortly before the assignment deadline. The sample participants reached an average of 64.57 points in the peer-reviewed assignment, out of 72 possible points.

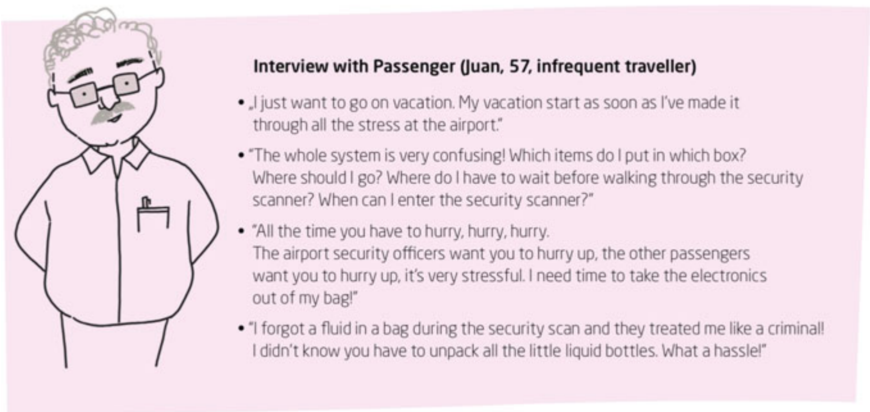


Fig. 2 Interview quotes of user "Juan" from the case material

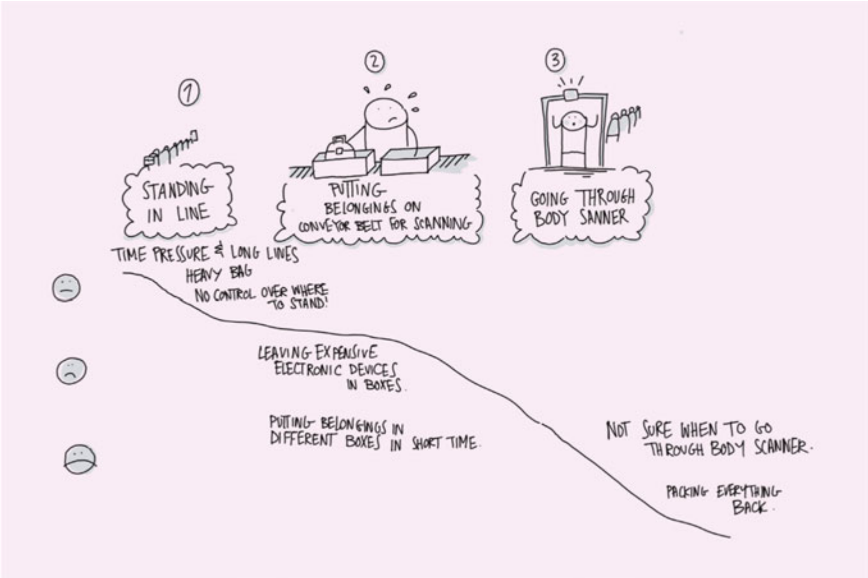


Fig. 3 User journey ("going through the security scan") example from the case material

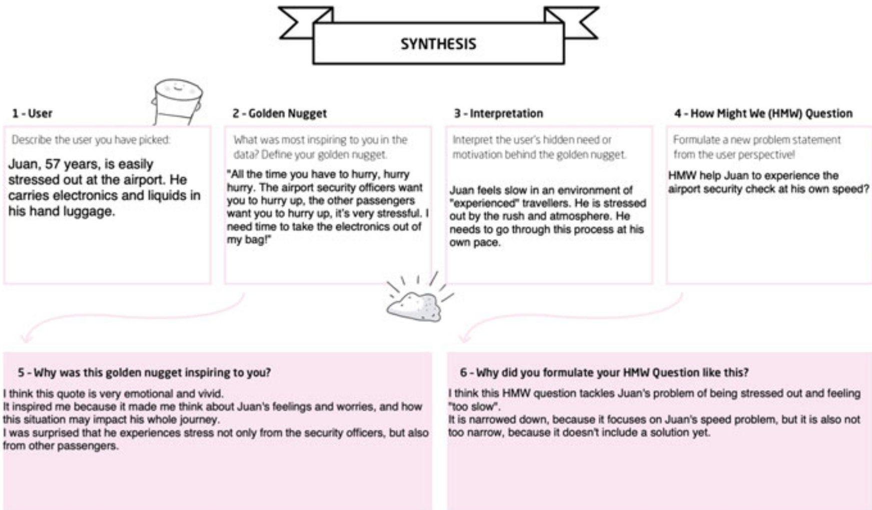


Fig. 4 Filled-in example template of the synthesis task

Table 2 HMW hashtags and higher order categories derived from all samples

HMW categories	HMW topics
Authority	Respect, authority, confidence, trust, recognition
Information	Information, clarity
Efficiency	Efficiency, quality time, speed
Simplicity	Simplification, self-help
Social	Stress, social, separation, interaction

Looking at data

All sample participants chose a user from the interview quote section of the case material. 11 went for an interview quote by "Lara", whereas five and four learners respectively chose "Juan" and "Alima." Only two participants chose "Michael."

Taking all samples into account, we identified a range of "How Might We Question" (abbr: HMW) topics. In the next step, we boiled the topics down to five categories: authority, information, efficiency, simplicity and social (see Table 2). These HMW categories depict different opportunity fields that participants identified for introducing a change through design.

The next step in the assignment was to facilitate a brainstorming session based on the new HMW question. Participants had to fill in an idea dashboard briefly describing their team's favored idea. The final ideas of our random sample were very diverse. Nonetheless, some topics recurred, independent of the user or the HMW category they chose. We therefore derived the following idea topic clusters: info, job, comfort and categories.[3]

Figure 5 portrays all synthesis paths from our sample.

It is interesting to see users triggered different participants to choose different user needs that they summarized in their HMW question. Also, we can see recurring HMW categories as well as idea categories for different users in the decision paths of participants. When working on a design challenge, it would be interesting to compare the different needs and HMW question topics that participants come up with among different users from the case material, and see if these needs match when the same HMW category that arose based on different users or if different users have different or contrasting needs within the same HMW category.

Possible implications

The preliminary analysis of the synthesis assignment task impelled us to think about the implications for design thinking teaching, both in physical and digital realms.

[3] *Info*: All ideas that aimed to add or change information material or display at the customs location.

Job: All ideas that were about making changes to the different jobs or job roles involved in this airport zone.

Comfort: All idea that played into increasing the comfort or decreasing discomfort of one or more involved users in the customs scenario (e.g. traveler, employee).

Categories: All ideas that introduced a different or new way to systematically label, order or categorize something (e.g. an object or a process).

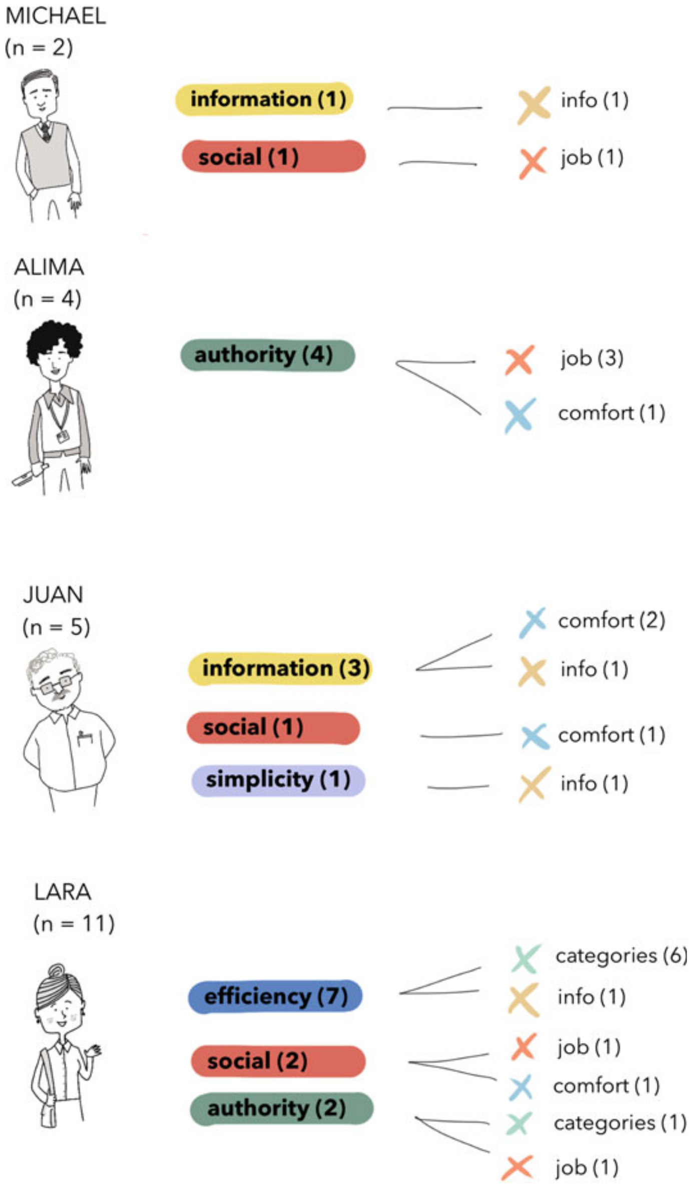


Fig. 5 Sample participants' synthesis paths, starting from selected users via HMW categories to idea categories. The numbers in brackets indicate how many participants came up with a HMW question or idea that we assigned to these categories

One of the common practices in teaching design synthesis to novices is to break down the phase into various small steps: sorting research data, identifying material, inferring insights and discovering hidden needs. Since synthesis is about converging, coaches often encourage teams to identify "golden nugget frames" from the research data. These are interesting or surprising aspects in the research data. After teams identify various nuggets, they try to infer insights or hidden needs behind these nuggets. Finally the team votes on the insight or the need they deem interesting and create a POV statement (or redefined design challenge) based on it. However, teams often have a difficult time with converging and leaving aspects behind, or deciding on one insight with which to move forward.

The common practice is to push teams to agree on a golden nugget (after each team member chooses their personal golden nugget and discusses why this nugget is interesting to them) and move forward from there as a team. Though, when faced with indecisive teams, some coaches encourage each team member to pick the insight they feel inspires them, create a POV statement individually and then present it to the rest of the team. This eases the decision process for some teams.

In relation to our observation of individual synthesis behavior in a MOOC, we are interested in further research on this coaching technique: when and how is it advisable to split teams up during a synthesis process? From our data, we see that various opportunity fields for redesign could be detected if team members first come up with individual synthesis paths.

We note that this is a practice that is sometimes implemented in educational settings already, but not necessarily in a structured way or following a scientific reasoning. Therefore, it would be worthy of further investigation.

All in all, our research shows that it is worth documenting the synthesis path in a design project. Thereby all possible paths are showcased. When a team decides on following a promising path that turns out to be out of scope for their company, partner, or their user group, they can still trace back how they got there and then take a different synthesis path forward.

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

References

Bailey, J. R., Chen, C. C., & Dou, S. G. (1997). Conceptions of self and performance-related feedback in the US, Japan and China. *Journal of International Business Studies, 28*(3), 605–625.

Bali, M. (2014). MOOC pedagogy: Gleaning good practice from existing MOOCs. *Journal of Online Learning and Teaching, 10*(1), 44.

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

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

Chita-Tegmark, M., Gravel, J. W., De Loures, M., Serpa, B., Domings, Y., & Rose, D. H. (2011). Using the universal design for learning framework to support culturally diverse learners. *Journal of Education, 192*(1), 17–22.

Conole, G. (2014). The 7Cs of learning design—A new approach to rethinking design practice. In *Proceedings of the 9th International Conference on Networked Learning* (pp. 502–509). Edinburgh: University of Edinburgh.

Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of teacher education, 53*(2), 106–116.

Ginsberg, M. B. (2005). Cultural diversity, motivation, and differentiation. *Theory into Practice, 44*(3), 218–225.

Kieran, L., & Anderson, C. (2018). Connecting universal design for learning with culturally responsive teaching. *Education and Urban Society, 1,* 15.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212–218.

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). *Computers & Education, 80,* 77–83.

Merrill, M. D. (2002a). A pebble in the pond model for instructional design. *Performance Improvement, 41*(7), 41–46.

Merrill, M. D. (2002b). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43–59.

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

Reflective Tools for Capturing and Improving Design Driven Creative Practice in Educational Environments

Adam Royalty, Helen Chen, Bernard Roth and Sheri Sheppard

Abstract Many educational institutions teach design thinking as a way to enhance student creativity. But does design thinking really promote a creative practice? By comparing a design thinking process to a creative process we argue that, when done well, design thinking does promote creative practice. Furthermore, we present both a student-centered tool and an instructor-centered tool that capture design driven creative practice.

1 Introduction

Design Thinking is an approach people and organizations take to solve ambiguous problems in creative ways (Kelley and Kelley 2013; von Thienen et al. 2017). In light of its growing popularity, it is important to ask, *does teaching design thinking actually promote creative practice? And if so, how can design driven creative practice be captured and measured?* These are the two primary questions we will address.

This chapter has four sections. The first is a basic overview of design thinking, including a brief history and how it is related to the field of design. That will clarify what aspects of creativity design thinking attempts to promote. The next section compares the design thinking process to the creative process. This includes investigating

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

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

B. Roth
e-mail: broth@stanford.edu

H. Chen · S. Sheppard
Department of Mechanical Engineering, 475 Via Ortega, Stanford, CA 94305, USA
e-mail: hlchen@stanford.edu

S. Sheppard
e-mail: sheppard@stanford.edu

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_4

the behaviors design thinking relies on. In the final two sections, we present two tools for measuring design driven creative practice: one focused on student outcomes, the other focused on academic settings.

2 What Is Design Thinking?

The phrase "design thinking" grew in popularity in large part due to Richard Buchanan's article *Wicked Problems in Design Thinking* (Buchanan 1992); though as Buchanan notes, much of this way of working was articulated by John Dewey in the early 20th century (Buchanan 1992). An earlier article, Designerly Ways of Knowing, outlined the value of problem solving like a designer (Cross 1982). In the article Cross argues that the creative ways that designers approach problems can be utilized by others. This suggests that design, like creativity (Scott et al. 2004), can be taught. In fact, it has been demonstrated that non-designers can learn and implement design (Royalty 2018). For this chapter we define design thinking as a practice where (primarily) non-designers use a design process with the goal of solving an open-ended challenge in a creative way.

It is clear that design is a creative field and that designers are generally creative individuals. Ray Eames, for example, is one of the 20th century's greatest creative luminaries (Kirkham 1998). Leonardo de Vinci was a spectacular designer and arguably the most prolific creative in human history. Although design does not have an agreed upon definition, one of the most popular is, *a course of action aimed at changing existing situations into preferred ones* (Newell and Simon 1972). Compare this to one definition of creativity, *a response or product determined to be both original and relevant* (Runco and Jaeger 2012; Stein 1953). "Seeking preferred situations" is nearly identical to a response or product as being "relevant"—sometimes alternatively described as useful. "Changing an existing situation" implies seeking something "original." This means that design is creative by definition (at least by these common definitions).

If design is creative, then must it follow that design thinking is also creative? Unfortunately, it is not that straightforward. Design thinking is the practice of problem solving like a designer. It does not guarantee that the practitioner actually succeeds at designing. Because the practitioners are non-designers, they need assistance. Design thinking has a number of scaffolds to support successful implementation of design (Royalty et al. 2015). The design thinking process is the most common scaffold. Therefore, to determine if design thinking really promotes creativity, it is necessary to study how it is taught and applied by non-designers.

The Hasso Plattner Institute of Design at Stanford University (d.school) was one of the first groups in higher education to explicitly teach design thinking. Its methods are used around the world (von Thienen et al. 2017). The d.school has roots in industry; it shares a co-founder—David Kelley—with the leading global design thinking firm IDEO. As a result, the design process taught at the d.school is virtually identical to a leading design process used in industry. This paper examines the design thinking process taught at the d.school for our analysis. The next step is to determine what aspects of creativity to use in this comparison.

The d.school was founded, in part, to *develop a sense of creative confidence in students* (Kelley and Kelley 2013). Courses use a design thinking process and experiential learning to help students solve real world problems, often with project partners (Dym et al. 2005; Mitroff et al. 2013; Royalty 2018; von Thienen et al. 2017). d.school students have produced a number of highly creative outcomes including a low-cost infant warmer for the developing world, an innovative app purchased by a major tech company, and a redesign of the pediatric MRI scanning experience (Kelley and Kelley 2013). However, the teaching goal of the d.school is not to create world-changing products but rather to help people solve problems creatively in their lives.

Using 4-C model of creativity developed by Kaufman and Behetto, it is clear that the intent of design thinking is to impart little c or mini-c achievements (Kaufman and Beghetto 2009). The idea is that the "little" creative accomplishments will help people be better, more creative problem solvers over the course of time. There is evidence that d.school alumni do demonstrate little c and mini-c accomplishments (Royalty et al. 2012). Alternatively, some might argument is that organizations should employ design thinking to bring about radical change, which would fit into the Pro C or Big C category. Although design thinking can help organizations innovate (Brown 2009; Royalty and Roth 2016), the 4-C model of creativity applies to individuals. This chapter focuses on the extent to which design thinking promotes creative practice in a person.

To summarize Sect. 1, the design thinking process codified at the d.school is what this chapter will compare to a creative process. In addition, the outcome of little c and mini-c achievements inform the measurement goal discussed in Sect. 3.

3 The Design Process Versus the Creative Process

The design thinking process taught at the d.school (Fig. 1) has five steps; Empathize, Define, Ideate, Prototype, and Test. Although the process does not have to be followed linearly, most courses, workshops, and projects in industry tend to follow the steps (von Thienen et al. 2017). Each step employs a different set of tools and dispositions. Tools refer to specific types of actions or activities like low resolution prototyping, and dispositions are how people approach work within a specific step. In this way dispositions are reminiscent of de Bono's six thinking hats (de Bono 1995). However, while solving problems through design, one is much more likely to explicitly identify the process steps than the disposition. Some dispositions, like always keeping user needs at the forefront (human-centered) tend to permeate the entire process (Goldman et al. 2012). Other dispositions like rapid idea generation sifting through ambiguity are specific to one or two steps.

The creative process used for comparison is Preparation, Incubation, Illumination, and Verification (Wallas 1926). We used the four phase creative processes over

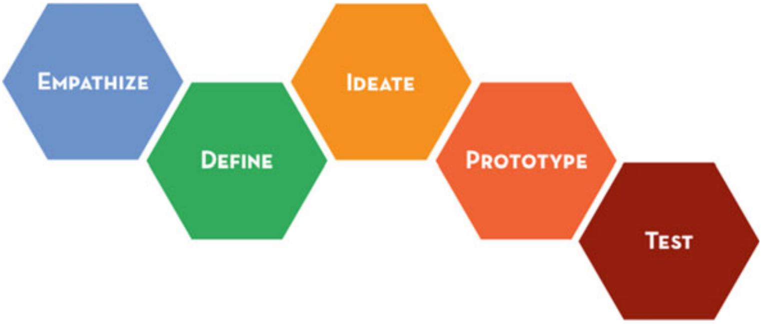


Fig. 1 Stanford d.school design thinking process

a similar five phase creative processes which ends with Elaboration–where the idea is implemented. The design thinking process leads to a nearly identical implementation phase (Buchanan 1992; Cross 1982; von Thienen et al. 2017). Therefore, we did not feel it was necessary to include Elaboration. The analysis presented in this chapter uses the design thinking process as a baseline (as detailed below), then draws connections to the creative process.

Empathy

Design thinking typically addresses challenges that are open-ended and ambiguous, meaning there is not a clear direction or deliverable. The challenges frequently involve people—customers or users. The first step of the design thinking process—empathize—encourages exploration of the problem space, particularly by understanding how the challenge impacts people (Mitroff Silvers et al. 2013). It is a time of intense data gathering through tools like interviewing, observation, and secondary research (Mitroff Silvers et al. 2013). The dispositions this process step encourages are curiosity, openness, and empathy (von Thienen et al. 2017). The outcome of the empathize step is a large amount of unstructured data. This includes quotes, interview transcripts, photos, sketches, internet reports, and more.

Define

Define is the second step of the design thinking process. The data collected during the empathize step are organized, categorized, and sorted. There is not a prescribed way to do this. There are several different tools available including 2 × 2 grids, user empathy maps, and POV statements (Mitroff Silvers et al. 2013), and the actual data collected suggest what tools to use. The dispositions required are slightly different from the empathize step. Integrative thinking, associative thinking, and Janusian thinking (Rothenberg 1971) help practitioners make sense of the data. The goal of the define step is to sift through the ambiguity of the challenge and create a clear problem objective to be solved. This often happens by identifying one or two critical needs of the people affected by the challenge (Mitroff Silvers et al. 2013; von Thienen et al. 2017).

Ideate

Idea generation happens in the third step, ideate. Participants generate multiple ideas to address the problem objectives articulated during the define step. The primary tool used for this is brainstorming (Osborn 1953; Mitroff Silvers et al. 2013). However, variants like bodystorming and recombinant generation also exist. Ideate requires open, unfiltered, and energetic dispostions. The goal is to come up with as many ideas a possible—regardless of feasibility or viability—then select one to three to move to the next step (von Thienen et al. 2017).

Prototype

The prototype step is where the ideas are first created physically. That is not to say that all the ideas have to be products, as services and experiences can also be prototyped (Mitroff Silvers et al. 2013). The actual construction usually begins by creating a few low-resolution versions (von Thienen et al. 2017). These might be made in 15 min or less out of materials like paper, tape, and post-its. The dispositions associated with this process step are inventiveness, openness, and resourcefulness. It is important to note that construction skills, like those professional designers or engineers have, are useful but not required. In design thinking initial prototypes can be made regardless of technical ability (von Thienen et al. 2017). At the end of the prototype step, participants have one or two prototypes ready to test.

Test

The final step is test. The prototype or prototypes are tested with users and iterated upon. There are different types of tests; A|B testing, usability testing, and experiential testing. For each type, the goal is to learn what aspects of the prototype work and what aspects do not. Analytic thinking and synthetic thinking are necessary for the test step. After this step is complete the design process cycle may begin again to refine the concept. Alternatively, if the idea is ready, it may be implemented.

Each element of the creative process Wallas described has its own purpose. All the constraints and relevant problem information are collected during the preparation phase. During the incubation phase a person consciously and unconsciously processes the problem while searching for solutions. The solution unveils itself during illumination. Finally, the appropriateness of the solution is explored as part of the verification stage.

So, how do these two processes compare?

Empathize appears to overlap significantly with preparation. Both are about collecting relevant information. It is important to highlight that empathize tools tend to focus on people and understanding their needs. The preparation phase does not dictate how one collects data—anything goes. Therefore, empathize might be a type of preparation.

Define, like empathize, overlaps with preparation, although this part of the design thinking process is not about collecting data. It is about organizing data in a way to prepare participants to solve a problem. Define may correspond with the activities one performs in the latter part of the preparation phase. This suggests that empathize corresponds with activities at the beginning of the preparation phase.

Ideation partially aligns with incubation. The idea generation tools in design thinking where participants explicitly generate new ideas connects with the conscious work theory (Runco 2014). Furthermore, recombinant idea generation intentionally links ideas that are not obviously connected. This is similar to Synectics (Gordon 1961). The goal is to activate more remote associations, a very creative practice (Mednick 1968). However, design thinking does not actively promote unconscious work or, a related, recovery from fatigue theory (Runco 2014). This is not to say that a participant's subconscious does not work on the problem during design thinking, but the ideation phase typically happens quickly without much time to incubate internally. It is also the case that ideate is often when the solution arises, meaning that it aligns with illumination as well.

The combination of prototype and test relate to verification. A prototype is the manifestation of an idea created explicitly to test it. As with verification, the goal is to see if the solution actually solves the problem.

The analysis above suggests that the design thinking process is extremely similar to the creative process. Empathize and ideation each partially map to the creative process. It may be that the design thinking process is a subset of the creative process. It is also likely that some differences are more difficulty to detect. To gain more clarity, this chapter will perform a second comparison based on when each process leverages convergent and divergent thinking.

Another representation of the design thinking process is a flare/focus diagram (Fig. 2). It describes what parts of the design thinking process call for more ideas (divergent thinking) and what parts call for driving towards a clear goal (convergent thinking). This can be compared to when the same types of thinking happen in the creative process.

Through this comparison, the connection between empathize and preparation is not so straightforward. Empathize requires divergent thinking, whereas preparation requires convergent thinking. In this analysis define, and its associated convergent thinking, appear to be the stronger link. The large flare occurring during ideate does correspond with the divergent thinking in incubation and illumination. Prototype,

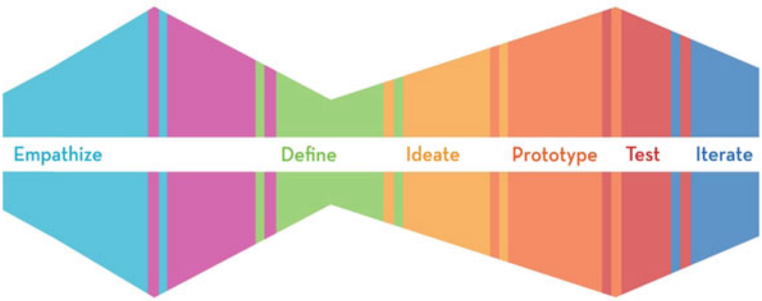


Fig. 2 The flare/focus design thinking process

which uses divergent thinking, appears to overlap less with the convergent thinking of validation. Test, however, is still a match.

This analysis suggests that prototype might connect more closely with incubation and illumination. One can argue that the act of prototyping is not simply representing a conceptualized idea (Beckman and Barry 2007). Making necessarily involves generating new ideas because physical constraints inspire improvisation and modification. One phrase often used to describe prototyping is “build to think” (Carroll et al. 2010). In this case, test would be the only design thinking step that connects with validation.

The way empathize relates to the creative process is now less clear. Is it a form of pre-preparation? Can it be that preparation involves *some* divergent thinking? Even though empathize is an exercise in divergent thinking, we argue that it is still a part of the preparation phase because the primary goal is to gather relevant data that will undergo a convergent thinking process during define.

Ultimately, neither the design thinking process nor the creative process are completely rigid (that wouldn’t be very creative!), which is another commonality they share. This means that it is not possible to map the processes on to each other in every context. However, there is a great deal of overlap which suggests that, from a process perspective, design thinking could promote creativity.

It is worth taking some time to comment about two behaviors the design process evokes and comparing them to known creative behaviors. One is the energetic behavior of brainstorming. When participants brainstorm the environment is very positive and active. This is reminiscent of a manic state which is linked to creativity (Andreasen 1997). The major difference being that the energetic state design thinking calls for is artificial and does not last nearly as long. The second behavior is openness to new ideas. This is a personality trait that predicts creativity (Puryear et al. 2017). In the first analysis of the design thinking process, openness was the most common behavior—found in three of the five steps. Although a complete behavioral analysis is needed, initial findings suggest that design thinking promotes at least two very powerful creative behaviors.

Based on the evidence above, this chapter concludes that design thinking, if practiced well, does promote creative practice. Moreover, the design thinking process may be a subset of the creative process. There are, however, ways for design thinking to improve how it promotes creativity. More support of unconscious idea generation is needed. Also needed is an emphasis on what innate personality traits might be conducive to strong design thinking practice. It will be interesting to see how design thinking evolves. Does it stay fixed in its approach to creativity, or does is grow to incorporate new creative practices?

4 Measuring Design Driven Creative Practice

4.1 Background

Having established that design thinking can lead to strong creative practice, we turn our attention to measuring that practice. This is an important question because the ability to improve instruction and better support creative design in organizations depends on measuring creativity.

Capturing and assessing creative work faces a number of challenges. Many researchers do not agree on what criteria of creativity to assess (Plucker and Makel 2010). Furthermore, most of the existing creativity assessments were developed to use in a controlled setting, often research experiments (Amabile 1982; Carson et al. 2005; Guilford and Merrifield 1960; Torrance 1988; Welsh and Barron 1959). This extends to many assessments of design thinking (Hawthorne et al. 2016; Royalty et al. 2014; Saggar et al. 2015). We wanted to capture student work in an ecologically valid environment. Furthermore, because both the design process and creative process involve several steps performed over time, it is important to observe how students repeatedly practice creative work during the duration of a d.school course. To this end we developed a tool called Reflective Design Practice (Royalty et al. 2018). RDP was tested with 19 students at the Stanford d.school over three quarters. The following sections describe the tool and the output students generate while using it.

4.2 Materials

Students completed a weekly reflection throughout the 10-week quarter. They were asked to take a photo of an artifact they created that week while doing work for a d.school course and respond to three to five prompts about how the artifact was created. An artifact could be a physical asset like a prototype or a whiteboard after a brainstorm. It could also be an experience like interviewing users or a team meeting. Students uploaded their photo and corresponding reflections onto a Google Slide Template (Fig. 3). Midway through the quarter each student participated in a semi-structured 45-min interview with one of the instructors. The questions were divided into four general categories (Table 1). During the interview students were asked to look back through their entries to provide concrete examples in response to the questions. In the last week of the quarter students shared their entire set of reflections in small groups.

Fig. 3 Reflective design practice weekly reflection template

Table 1 Reflective design practice mid-quarter interview protocol

Category	**Environment—understanding how space, instructors, peers, and time constraints affect a creative practice**
Questions	What aspects of the environment supported the creation of the artifact?
	What aspects of the environment were **barriers** to the creation of this artifact? Why did they act as barriers?
Category	**Contrast—understanding how a creative practice differs from other working practices a student engages in prior to and after a design-based curricular experience**
Questions	How is the artifact different from what you might normally create in another (non-d.school) course?
	How did the d.school style of working enable (or not) the creation of the artifacts?
	How does this different from the style of working non-d.school courses enable?
	How would you approach integrating this process into one of your non-d.school courses?
Category	**Personal comfort/discomfort—understanding what parts of the design-based curriculum feel personally comfortable or uncomfortable to a student and why**
Questions	Which artifact was created using a style or way of working that felt comfortable or familiar to you?
	Which artifact was created using a style or way of working that felt the least comfortable or familiar to you?
Category	**Themes—noticing how certain themes appeared across multiple artifacts**
Themes	Comfort/discomfort with ambiguity
	Rapid prototyping
	Intangible behavior

4.3 Output

Over the course of the quarter, a student completes 10 Google Slide Templates. Each slide captures an example of actual creative work. The work can come from any part of the design process. For example, one student captured the result of a team synthesis session where they defined user needs that they felt compelled to address. Another student shared a prototype her team created in a campus dining hall aimed at helping students make more informed nutritional choices. The accompanying reflection describes how and why the artifact was made. All together the slides present a perspective on a student's journey through a d.school course. The mid-quarter interviews help students and instructors key in on areas of struggle and growth.

4.4 Conclusion

RDP is a flexible tool. Reflections can be assigned multiple times a week or scattered throughout a term. The prompts accompanying each photo can be modified to focus on a particular topic. What is important is that students capture real work and think deeply about how they created it. Furthermore, it is essential that students spend time reflecting across multiple entries. The ultimate goal is for students to develop insight into their own creative practice by observing it grow over time (Royalty et al. 2018). There is strong evidence that reflections prepares students to better transfer learning from an academic context to a real world context (Bransford and Schwartz 1999; Flavell 1979; Greeno et al. 1993).

The output of RDP can help instructors better understand how students experience a course as a whole. Because the reflection documents an entire learning journey, it can complement other reflections focused on particular techniques, class sessions, or projects. RDP can be useful to researchers, as they can code the interviews and slides using a number of different frameworks.

5 The Influence of Academic Settings on Design Driven Creative Practice

5.1 Background

The previous section illustrated a tool designed to capture student work. This section outlines another tool that instructors can use to describe their pedagogical approach to supporting design driven creative work. Building off of last year's work (Royalty et al. 2019) we sought to understand the how design instructors manipulated a learning context to better develop design practice within their students. We iterated on data gathering techniques used in previous studies (Royalty et al. 2014, 2018) to create a new tool for capturing how instructors teach. The tool has three general parts; perceived student journey, connectedness across campus, and environmental variables. Instructors begin by mapping out what and how they believe students learn design. Then they focus on the variables of the environment that most impact that journey.

We surveyed 27 instructors with experience teaching design-based curriculum. They were all participants at a conference of design educators. The instructors represented 13 universities and 5 colleges. Seventeen of the 18 institutions are located in the United States, with the lone international school being in Mexico. The median amount of experience teaching design was seven years.

We chose this group to study because we wanted to collect data from people committed to teaching design-based curriculum. At the same time, we wanted to include a wide range of academic contexts into our sample. Had we surveyed 27 instructors at the Stanford d.school, there is a good chance their responses would have been similar. Instead we wanted to be expansive and leverage the continually growing network of educators.

5.2 Materials

Our goal was to understand how context impacts the journey instructors take students on. That entails understanding—from the instructor's point of view—where students are at the beginning of the journey, where they are at the end, and what happens along the way. We asked the questions in that order because we wanted them to focus on learning outcomes first, then map those to how the students actually achieved those outcomes.

It is important to note that a student's journey learning design is not contained in the interaction with one teacher. So, when we say beginning and end we mean the beginning and end of the time the instructor has direct influence on that student—most often through teaching a course. We decided to scope this so that we could more easily compare responses across instructors and courses. Additionally, we asked instructors to think of a particular student rather than a generalization of students so that they could add specific details. They were given the option of mapping the journey of two students.

Figure 4 shows the worksheet instructors use to articulate where they believe students are before and after a design-based learning experience. This includes expressing what the learning goals and the emotional goals of instruction are. In both the before and after case instructors identify students' conception of design and creativity. They also list students' approach to problem solving and learning.

A journey map, Fig. 5, illustrates how instructors move students from the before state to the after state. We asked them to capture any design-based learning experience that a student might engage in at their institution. This could include courses,

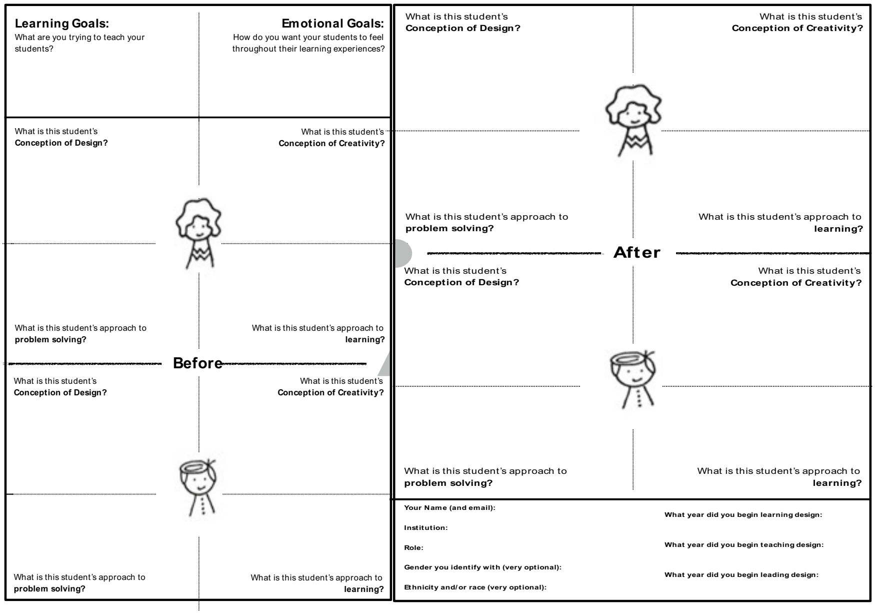


Fig. 4 Worksheets capturing instructors' perception of how students conceive of design, creativity, and problem solving before and after a design-based learning experience

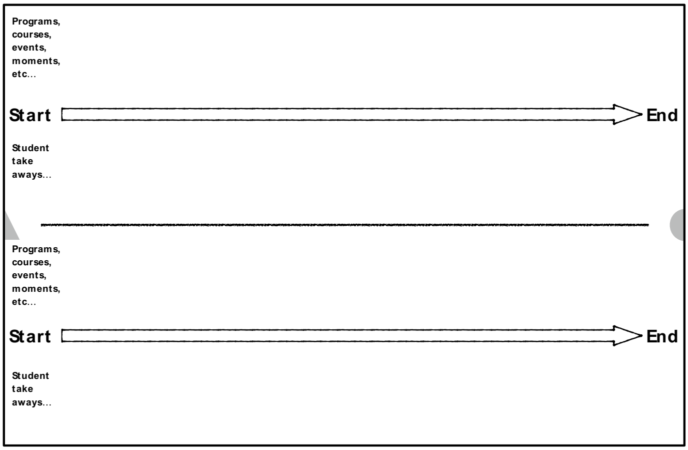


Fig. 5 A journey where instructors listed the different learning experiences students typical engage in; as well as, a desired student take away

workshops, coaching, etc. With each entry, instructors also wrote what key lessons they intend student to take away.

The second exercise we asked instructors to complete was an ecology illustrating how their institute, department, or course connects with the rest of the university (Fig. 6). This gives us data on the type of design efforts in different institutions. We can then compare contexts across institutions—and how those institutions relate to the rest of their university. For example, the environment in an institute like the d.school is different from a place that has a single design course without a dedicated space.

Finally, we asked instructors to list what variables they control while creating a design experience, see Fig. 7. They were given five minutes to list as many variables they play with, account for, or otherwise design into their learning experiences. The full paper (Royalty 2018) has more details but we will summarize the findings in the results section.

In the end, each instructor created a single Design Practice Canvas (Fig. 8) connecting students' learning journey and the variables used to support instruction. This proved to be a useful tool for the instructors to reflect upon the work they do, much of it implicitly, to develop students' design practice. This suggests that the canvas could be extended for use in non-academic contexts to help leaders better create environments that support design work.

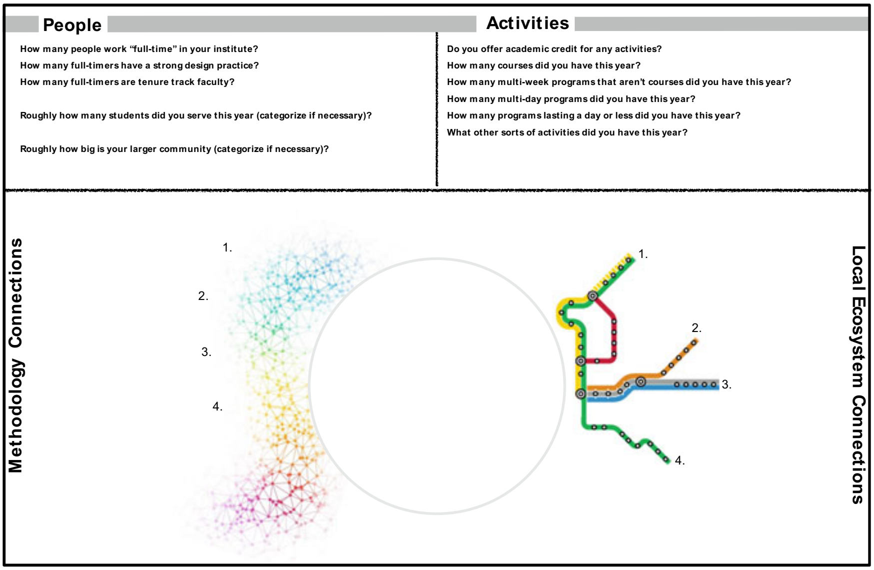


Fig. 6 Design ecology illustrating connections between design-based experiences and the university at large

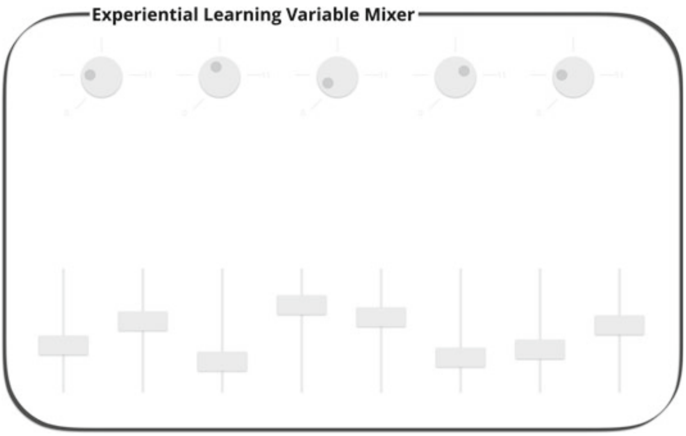


Fig. 7 Worksheet capturing variables design instructors use to create a design-based learning experience

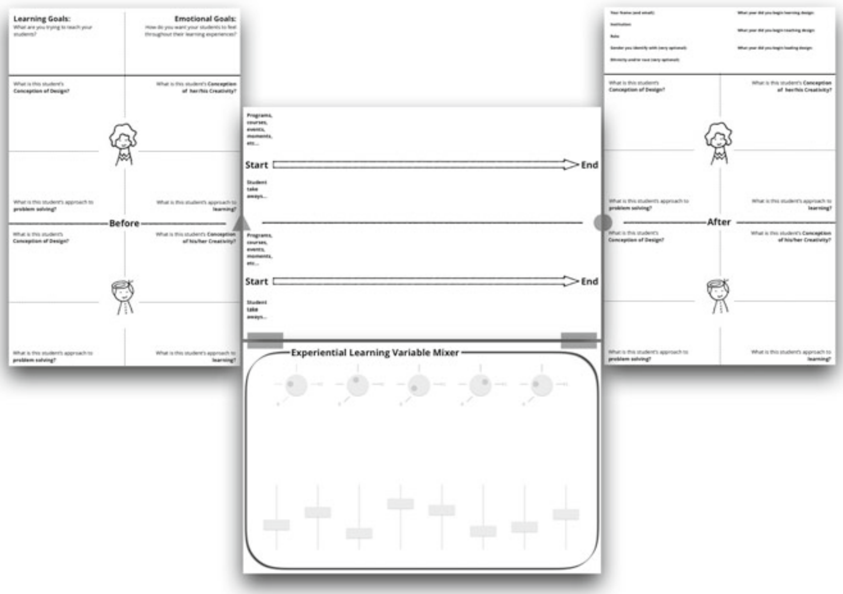


Fig. 8 Design practice canvas

5.3 Output

We are still in the process of analyzing the Design Practice Canvas output. A major aspect of the analysis focuses on the creative dispositions instructors try to teach their students. We will code the before and after sections of the journey maps and look for patterns between creative dispositions and the type of institution. One potential finding is whether or not environments that support prolonged student engagements (e.g. full semester courses) seek to teach different creative dispositions than environments that focus on short term engagement like workshops. Another question is what are the most common activities instructors employ to increase design driven creative practice.

The design ecologies will be analyzed for patterns across different institutions. It is already clear that no two institutions teach design-based curriculum the same way. A cursory review shows that some places offer multiple courses, others a single design-based course, and a few offer no credit bearing experiences. This is not a measure of success or impact. It simply helps the community understand the current diversity of approaches.

5.4 Conclusion

The Design Practice Canvas was created to help an institution articulate its approach towards teaching and how it exists relative to the larger college or university. This can help instructors be more intentional about how they evolve design-based curriculum. It also helps them understand what resources and partnerships the can seek within their broader ecosystem. Beyond these two benefits, a larger goal is to help instructors better learn from one another by understanding the varied contexts in which people practice design-based pedagogy. Ultimately our hope is that we can provide different models for organizations seeking to implement design-based pedagogy.

6 Conclusion

This chapter began with a definition and brief history of design thinking. We then compared a design process to the creative process by mapping two different representations of the design process used at the d.school to a well-known creative process. The analysis determined that design thinking does promote creative practice. Finally, we shared two different tools for understanding design driven creative practice. Analysis of the data collected already through these tools should give us further insight into how people teach and learn design. Moreover, we designed the tools to be useful for instructors looking to improve their own teaching practice.

References

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. *Journal of Personality and Social Psychology, 43*(5), 997.

Andreasen, N. C. (1997). Creativity and mental illness: Prevalence rates in writers and their first-degree relatives. *Eminent Creativity, Everyday Creativity, and Health*, 7–18.

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

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

Brown, T. (2009). Change by design.

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

Carroll, M., Goldman, S., Britos, L., Koh, J., Royalty, A., & Hornstein, M. (2010). Destination, imagination, and the fires within: Design thinking in a middle school classroom. *International Journal of Art and Design Education, 29*(1), 37–53.

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal, 17,* 37–50.

Cross, N. (1982). Designerly ways of knowing. *Design Studies, 3*(4), 221–227.

de Bono, E. (1995). Serious creativity. *The Journal for Quality and Participation, 18*(5), 12.

Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education, 94*(1), 103–120.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist, 34*(10), 906.

Goldman, S., Carroll, M. P., Kabayadondo, Z., Cavagnaro, L. B., Royalty, A. W., Roth, B. … Kim, J. (2012). Assessing d.learning: Capturing the journey of becoming a design thinker. In *Design thinking research* (pp. 13–33). Berlin, Heidelberg: Springer.

Gordon, W. J. (1961). Synectics: The development of creative capacity.

Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). *Transfer of situated learning*.

Guilford, J. P., & Merrifield, P. R. (1960). Structure of intellect model: Its uses and applications. University of Southern California.

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

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

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

Kirkham, P. (1998). *Charles and Ray Eames: designers of the twentieth century*. mit Press.

Mednick, S. A. (1968). The remote associates test. *The Journal of Creative Behavior, 2*(3), 213–214.

Mitroff Silvers, D., Rogers, M., & Wilson, M. (2013). Design thinking for visitor engagement: Tackling one museum's big challenge through human-centered design. In *Museums and the Web*.

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

Osborn, A. F. (1953). Applied imagination, principles and procedures of creative thinking.

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. *The Cambridge handbook of creativity* (pp. 48–73).

Puryear, J. S., Kettler, T., & Rinn, A. N. (2017). Relationships of personality to differential conceptions of creativity: A systematic review. *Psychology of Aesthetics, Creativity, and the Arts, 11*(1), 59.

Rothenberg, A. (1971). The process of Janusian thinking in creativity. *Archives of General Psychiatry, 24*(3), 195–205.

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

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

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

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

Royalty, A. (2018). Design-based Pedagogy: Investigating an emerging approach to teaching design to non-designers. *Mechanism and Machine Theory, 125*, 137–145.

Royalty, A., Chen, H. L., & Sheppard, S. (2018). Reflective design practice: A novel assessment of the impact of design-based courses on students. In *Proceedings of the ASEE/IEEE Frontiers in Education Conference*, San Jose, CA.

Royalty, A., Chen, H., Roth, B., & Sheppard, S. (2019). Measuring design thinking practice in context. In *Design thinking research* (pp. 61–73). Cham: Springer.

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092.

Runco, M. A. (2014). *Creativity: Theories and themes: Research, development, and practice*. Amsterdam: Elsevier.

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

Scott, G., Leritz, L. E., & Mumford, M. D. (2004). The effectiveness of creativity training: A quantitative review. *Creativity Research Journal, 16*(4), 361–388.

Stein, M. I. (1953). Creativity and culture. *The Journal of Psychology, 36*(2), 311–322.

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

von Thienen, J., Royalty, A., & Meinel, C. (2017). Design thinking in higher education: How students become dedicated creative problem solvers. In *Handbook of research on creative problem-solving skill development in higher education* (pp. 306–328). IGI Global.

Wallas, G. (1926). The art of thought.

Welsh, G. S., & Barron, F. (1959). *Barron-Welsh art scale: a portion of the Welsh figure preference test*. Psychologists Press.

Augmenting Learning of Design Teamwork Using Immersive Virtual Reality

Neeraj Sonalkar, Ade Mabogunje, Mark Miller, Jeremy Bailenson and Larry Leifer

Abstract When it is done well, design teamwork is a fun, creative, and productive activity. However, the learning of effective design teamwork is hampered by lack of exposure to variation in design contexts, lack of deliberate practice and lack of appropriate feedback channels. In this chapter we present immersive Virtual Reality (VR) in accompaniment to action-reflection pedagogy as a solution to augmenting design team learning. A prospective case of using VR to augment design teamwork practice is discussed and a research agenda is outlined towards understanding VR as a medium for design teamwork, investigating its influence on design team self-efficacy and implementing it in design education courses.

1 Introduction

Multidisciplinary teamwork is a key value in design thinking (Dym et al. 2005). Effective design teamwork can be defined as the performance of interpersonal interactions in such a manner that design tasks are collaboratively completed with the result being greater than the sum of individual contributions. This occurs through each individual supporting, challenging and building on each other individual to deepen user insights, generate novel problem frames, develop bold concepts, and

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

A. Mabogunje
e-mail: ade@stanford.edu

L. Leifer
e-mail: leifer@stanford.edu

M. Miller · J. Bailenson
Virtual Human Interaction Lab, Stanford University, Stanford, CA 94305-2050, USA
e-mail: mrmillr@stanford.edu

J. Bailenson
e-mail: bailenso@stanford.edu

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_5

prototype effective solutions—ultimately resulting in an emergent design outcome that is beyond any individual imagination and capability. While this ability to work effectively as a design team is prized in the professional world, the learning of design teamwork in a design education is afforded a secondary importance. Students learn teamwork by participating in one or more design projects where the instruction is focused primarily on design process models. Team issues are dealt with as and when they emerge. This has the consequence of some students getting to work in mediocre or a priori effective teams and not learning to master the challenges of design teamwork, and some students getting to work in ineffective teams without having proper mechanisms for overcoming their ineffectiveness and developing team self-efficacy. These students often give up on design teamwork after having particularly painful team experiences. Is there a way to address this shortcoming by creating learning experiences for students to master design teamwork and build team self-efficacy?

In this chapter, we outline the key challenges we face as teachers and students in learning design teamwork, and propose immersive Virtual Reality as a technological accompaniment to action-reflection pedagogy to augment the learning of effective design teamwork.

2 Challenges to Design Team Learning

There are three key challenges to learning effective teamwork in design courses.

2.1 Lack of Exposure to Varying Context

Design is a context-dependent activity. The individual, organizational, task and environmental context in which a design team activity occurs influences the effectiveness of that activity. For example, when a design team meets to generate product solution concepts, the level of interpersonal hierarchy in the team, the team members energy levels and motivations, the nature of the design challenge—whether it's a consumer product or a systems level problem, the physical environment in which the team is meeting, all of these influence the interpersonal interactions that form design teamwork and make this team situation different from other concept generation sessions that same team may have had in the past. In spite of this, when we currently teach design, we teach it through a single design project or at best a few different short projects in a course setting. These hardly capture the contextual variability that a student would need to master in order to prepare for effective design teamwork in a professional setting. The key challenge here for design instructors is to comprehensively capture the context variables for design team activities and then create varying design situations for students to practicing their team performance.

2.2 Lack of Deliberate Practice

The project-based learning (PBL) pedagogy often used by design courses is helpful in facilitating active learning in students by giving them a realistic project situation. However, these PBL projects often are one-off experiences that do not afford students opportunities to practice and master design teamwork. Ericsson et al. (1993) in their seminal paper on acquisition of expert performance presented deliberate practice—a regimen of effortful activities targeted towards improving performance and overcoming motivational and external constraints—as an underlying determinant for developing expertise. Developing design teamwork effectiveness requires this deliberate practice more than mere exposure to team situations that is currently prevalent in design courses. This presents a unique challenge to design instructors—how to create an arena for students to practice design teamwork in a deliberate manner?

2.3 Lack of Appropriate Feedback Channels

The third challenge facing design team learning is the lack of feedback channels for improving design teamwork. For example, when a conflict situation arises in a team, the team members' own emotional maturity and social skills are relied on more than external feedback as to how the individuals are doing in their handling of the situation. The instructors at times may give feedback, but more often it is observed that instructors focus on product and process level feedback, while students are left to handle interpersonal team interactions, and interpersonal dialogue and feelings on their own. This is a function of instructors facing a shortage in time for coaching, and a lack of feedback channels and tools available to students to address teamwork issues. The result is that even if potential learning situation do occur in design teams, students are ill prepared to take advantage of them and learn through receiving and acting on appropriate feedback. As we address the first two challenges for design team learning—exposure to multiple contexts and creation of an arena for deliberate practice, the availability of feedback becomes the next important limiting factor. We propose that feedback needs to be built-into the practice arena for design teamwork such that students can self-correct and generate new interaction behaviors all by themselves as they keep on practicing.

3 Virtual Reality as a Medium to Augment Design Team Learning

As instructors and students of design teamwork, we ourselves grappled with the three challenges presented in Sect. 2. Our prior work dealt with the creation of a pedagogical framework that involved creation of multiple situational learning experiences followed by embodied reflection (Mabogunje et al. 2018). We were still missing a way to create a distinct arena for deliberate practice along with the capability to vary design context parameters at low cost. While looking for solutions to overcome this challenge, we found immersive Virtual Reality to be a promising technological solution.

Immersive Virtual Reality (VR) refers to the experience of being immersed in a virtual world through the use of head-mounted display and sensors that track head and body movement in space. This is an embodied experience that is different from the sensory experience of watching a virtual environment on a 2-D or even a 3-D screen. The ability to move your head and your body and access the different perspectives in a virtual world as if 'you are there' is key to achieving a simulation of reality in which you could practice design team behaviors (Bailenson 2018). See Fig. 1 for a visual of two designers wearing VR headsets and interacting in a virtual world.

VR with multiple users in a shared world has been studied for decades (Takemura and Kishino, 1992; Churchill and Snowdon 1998). One of the canonical use cases within this area of work is using virtual reality as a collaborative design tool (e.g., Leigh et al. 1996). The spatial nature of VR has lent itself to design tasks, especially in architecture (Rosenman et al. 2007). Furthermore, the unique aspects of mediated small-group behavior have been studied in virtual environments (Slater et al. 2000).

What is new in the current study is the simulation of the social situation a designer may find him or herself in. We propose creating virtual environments that mirror the various physical environments in which design teams might operate in, and then let students wear headsets to be immersed in that environment and interact as a team to achieve pre-specified design tasks. A variation of this VR arena for design teamwork might include pre-set characters and situations which ground the team in behaving in a certain way to practice dealing with specific design team situations. For example, we could create a conflict scenario involving a social loafer character who is not working as much as his team members and is actively shirking responsibility. Team members could then practice having an alignment conversation with this team member, who could be played by one of the students. By practicing such interactions over and over again, each time trying out something different, the students gain deliberate practice in handling difficult team situations. In more technologically advanced version of the VR teamwork arena, pre-set characters could be programmed virtual agents that

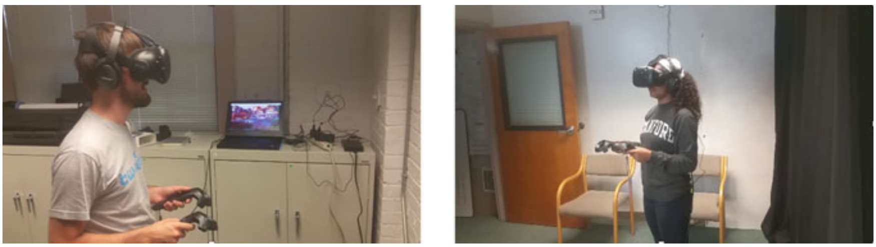


Fig. 1 Two designers wearing VR headsets and interacting in a virtual world

students could practice interacting with. VR could then enable students to overcome two of the limitations related to learning design teamwork – variation in contexts and availability of an arena for deliberate practice. The third limitation of the feedback channel could be addressed through embodied reflection tools coupled with team interaction analytics such as the Interaction Dynamics Notation (Sonalkar et al. 2013) that could be used to give meaningful feedback to teams practicing VR team simulations.

4 Affordances and Limitations of VR

While VR seems a promising medium for deliberate practice of design teamwork, it is important to be cognizant of its affordances and limitations as a design team simulation medium.

4.1 Affordances of VR for Simulating Design Teamwork

1. Immersion: The key affordance of VR that makes it particularly suitable for simulating design teamwork is that it is an immersive experience. Design teamwork is not a purely cognitive activity. It involves interacting with objects in the studio environment—markers, whiteboard, paper, prototypes etc.—and with people who are there in the space with you. VR can enable us to have this immersive experience of a design situation while working in a programmed virtual world.
2. Ability to change context variables: Multiple immersive VR environments could be programmed to match the variety of physical environments in which design teamwork activities occur in. Moreover, the avatars that people embody in VR could be changed to match the environments. For each of these environments, we could then create scenarios that outline the other context variables such as organizational hierarchy differences, motivational differences, task differences etc. The immersive nature of VR makes it easier to role-play these different variables than a real-life setting.
3. Repeatability of experience: Once a VR environment or scenario is programmed, it is available for repeated use. Students could interact multiple times with the scenario each time varying a certain element of their interaction to prototype new outcomes. This repeatability of experience is a key element to VR being suitable for deliberate practice.
4. Ease of use: Engaging with team members in VR is as easy as putting on a VR headset. The current level of consumer grade VR equipment is designed to be comfortably worn for an extended period of time and works with most laptops with a dedicated graphics card.
5. Low cost: The cost of using VR for deliberate practice of design teamwork requires the creation of multiple VR environments, team scenarios and the avail-

ability of physical space where a team can wear headsets and step into the virtual world. The creation of multiple VR environments for team interaction can be achieved through platforms such as High Fidelity, Sansar, VR Chat, Rec Room etc. and the programming is low cost. We envision design courses having a dedicated studio space for VR design teamwork in the future.

4.2 Limitations of VR for Simulating Design Teamwork

VR for simulating design teamwork has a few limitations that could be overcome with suitable scenario design.

1. 20 min time-limit: In general, it is advisable to keep activities to a 20-min time limit, per design activity scenario. It is not recommended to exceed this time limit since extended VR presence could disorient a person from their real-world situation. The physiological effects of very long-term VR use are still unknown.
2. Lack of facial expressions: Faces in VR do not currently transmit facial expression of the participants, since the head-mounted display impedes facial tracking sensors. We expect this situation to improve with time as the technology develops. In absence of facial expressions, emotions can be conveyed effectively using tone of voice. Thus, VR teamwork might need to emphasize the use of voice modulation to a greater extent than is usual in the physical world.
3. Caring for physical safety when in virtual world: When a participant is immersed in the virtual world, they do not have an awareness of the physical world they are moving in while wearing their headsets. Hence researchers have a ‘spotter’ available to each participant who ensures that the person wearing the headset doesn’t run into physical barriers while moving in the virtual world. This necessitates having one support person per participant while using VR.

5 Prospective Case: Using VR for Augmenting Design Team Learning

In order to further clarify the role of VR in augmenting design team learning, we present the following prospective case. A prospective case is similar to a case study except that it outlines a prospective scenario rather a scenario from the past.

John, Jill and Emily are team members in a senior design course at a university in the US. Along with the studio component in which they work on a company sponsored project, the course has a design teamwork lab component in which students get to practice their design teamwork. John, Jill and Emily are going to attend their first lab session.

John is the last to arrive at the lab, which is a dedicated space for team VR. The lab instructor welcomes the team and guides each team member to their individual VR stations. Each individual gets a small 2 m by 2 m space in which to move around when wearing a headset. John is excited to try out VR. He has heard a lot about it and has even played a few games with his friend's VR set. Jill is a bit nervous. She is a bit wary of the new technology, but keen to see what it would feel like to work in a team in VR. Emily is quite eager to get on with the activity. She believes that she is not a great team player and wants to do all she can to improve her teamwork.

The instructor assists them with wearing the headsets and takes them to a virtual tutorial room where they get to select their avatar, look themselves in the mirror and play catch with a virtual ball. Emily and John take to the VR world with ease. Jill takes some time to figure out how to operate in the virtual world, but soon joins John and Emily in playing catch and starts having fun. After 5 min, the instructor guides them to return to the real world and take off the headset. He then hands each person their first design task. It is a concept generation task in which they are embodying the role of startup founders in a suburban garage developing a personal mobility solution. The team wears their headsets and now they are transported to a virtual garage complete with a Toyota in the driveway! The garage has a small desk and a whiteboard with markers. The team has 10 min to complete their concept generation task. Emily initiates the discussion and soon the entire team is busy conversing and discussing solutions for personal mobility. John gets up (in virtual world) and goes to the whiteboard and starts drawing a sketch. Emily and Jill join him in the sketching. 10 min go past quickly, and the instructor has to call them twice to step back out in the physical world. When they step out, the instructor hands them a personal reflection sheet for embodied reflection which they fill out in a couple of minutes. Next, they are given their interaction analytics feedback computed with the Interaction Dynamics Notation and they discuss how they performed in terms of supporting, building on and deepening concepts in their discussion. John realizes he needs to be more supportive of his team members rather than pushing his own ideas.

After a 10 min debrief, it is time for the next task. The team now gets a boardroom decision making task. They are part of a corporate design team that has a meeting with their chief product officer to determine which of three promising product lines to pursue. John, Emily and Jill don their headsets again and quickly become immersed in the world of corporate design decision making. This is followed by four more design scenarios and associated post-activity debriefs. Over a period of two hours, the team has practiced six different design scenarios and greatly broadened their teamwork repertoire. As John, Emily and Jill head home, Jill comments that this was the most intense and productive lab session she has ever been to at her university.

6 Research Agenda to Realize the Promise of VR

What are the research questions that need to be answered in order for us to realize the prospective case discussed above? In this section we discuss the research agenda that needs to be implemented to realize the promise of VR for augmenting design team learning.

The research agenda for investigating the use of VR for design team learning can be divided into three sections.

6.1 Research Targeting the Fidelity of VR Teamwork to Real Teamwork

An important requirement for VR to be an effective medium for deliberate practice of design teamwork is that it is able to render design team activity with sufficient fidelity to real teamwork. Our early exploration of VR as a teamwork medium are promising for design activities such as concept generation and decision making. However, the following research questions need to be investigated systematically before proceeding with VR.

What elements of design team activity translate to virtual world and what elements do not?
What characteristics of interpersonal interactions differ in a VR setting vs real world setting?

6.2 Research Targeting the Influence of VR Team Work on Participant Learning

Once we establish that design teamwork can be rendered in VR with sufficient fidelity, we need to investigate the influence on participating in VR design teams on participant learning. The following questions are relevant in this research.

What is the influence of VR team work on participants' team self-efficacy when compared to influence of similar teamwork in real world and its influence on participant team self-efficacy?
Do behavioral changes adopted in the virtual world in the domain of design teamwork translate to behavioral changes in real world design teamwork?
What is the frequency of exposure to different design scenarios needed to build design team self-efficacy?
What is the variety of scenario exposure needed to build design team self-efficacy?
What individual participant characteristics influence the building of design team self-efficacy using VR?
What are the feedback parameters that need to be included post-VR activity to build participants' design team self-efficacy?

Note that the key measurement of participant learning is team self-efficacy. Following Bandura's Social Cognitive Theory (Bandura 2001), self-efficacy is an effective predictor of future performance.

6.3 Research Targeting the Conditions for Designing an Optimal VR Experience for Design Team Learning

The third section of the research agenda pertains to building an effective VR practice arena for implementing the role of VR for deliberate practice in design courses.

How might design instructors engage with VR within the design course frameworks?
What is the optimal team size for maximizing the building of design team self-efficacy in VR?
What is the influence of having physically distributed team members participating in the same virtual team on the building of design team self-efficacy?

The question above pertains to the use of VR for distributed education as an immersive alternative to MOOCs.

7 Conclusion

With advances in technology and the availability of low-cost headsets, VR is becoming a promising technological medium to augment design team learning through deliberate practice and reflection. In this chapter we highlighted the key challenges facing design team learning and proposed the use of a VR based arena for augmenting design team learning. We outlined a research agenda for investigating and realizing the promise of VR. We have started implementing this agenda and hope that other researchers interested in design teamwork will take up this agenda so that collectively we can achieve the goal of helping students learn the art and science of effective design teamwork.

References

Bailenson, J. (2018). *Experience on demand: What virtual reality is, how it works, and what it can do*. WW Norton & Company.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology, 52*(1), 1–26.

Churchill, E. F., & Snowdon, D. (1998). Collaborative virtual environments: An introductory review of issues and systems. *Virtual Reality, 3*(1), 3–15.

Dym, C. L., Agogino, A. M., Eris, O., Frey, D. D., & Leifer, L. J. (2005). Engineering design thinking, teaching, and learning. *Journal of Engineering Education, 94*(1), 103–120.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363.

Leigh, J., Johnson, A. E., Vasilakis, C. A., & DeFanti, T. A. (1996). Multi-perspective collaborative design in persistent networked virtual environments. In *Proceedings of VRAIS '96* (pp. 253–260). IEEE. https://doi.org/10.1109/vrais.1996.490535.

Mabogunje, A., Sonalkar, N., & Leifer, L. (2018). What if we have become trees? *International Journal of Engineering Education, 34*(2), 686–694.

Rosenman, M. A., Smith, G., Maher, M. L., Ding, L., & Marchant, D. (2007). Multidisciplinary collaborative design in virtual environments. *Automation in Construction, 16*(1), 37–44. https://doi.org/10.1016/j.autcon.2005.10.007.

Slater, M., Sadagic, A., Usoh, M., & Schroeder, R. (2000). Small-group behavior in a virtual and real environment. *Presence: Teleoperators and Virtual Environments, 9*(1), 37–51.

Sonalkar, N., Mabogunje, A., & Leifer, L. (2013). Developing a visual representation to characterize moment-to-moment concept generation in design teams. *International Journal of Design Creativity and Innovation, 1*(2), 93–108.

Takemura, H., & Kishino, F. (1992). Cooperative work environment using virtual workspace. In *Computer Supported Cooperative Work Proceedings* (pp. 226–232). ACM. https://doi.org/10.1007/978-4-431-68204-2_11.

Exploring Effective Team Interaction

Hive: Collective Design Through Network Rotation

Niloufar Salehi and Michael S. Bernstein

Abstract Collectives gather online around challenges they face, but frequently fail to envision shared outcomes to act on together. Prior work has developed systems for improving collective ideation and design by exposing people to each others' ideas and encouraging them to intermix those ideas. However, organizational behavior research has demonstrated that intermixing ideas does not result in meaningful engagement with those ideas. In this paper, we introduce a new class of collective design system that intermixes *people* instead of *ideas*: instead of receiving mere exposure to others' ideas, participants engage deeply with other members of the collective who represent those ideas, increasing engagement and influence. We thus present Hive: a system that organizes a collective into small teams, then intermixes people by rotating team membership over time. At a technical level, Hive must balance two competing forces: (1) networks are better at connecting diverse perspectives when network efficiency is high, but (2) moving people diminishes tie strength within teams. Hive balances these two needs through *network rotation*: an optimization algorithm that computes who should move where, and when. A controlled study compared network rotation to alternative rotation systems which maximize only tie strength or network efficiency, finding that network rotation produced higher-rated proposals. Hive has been deployed by Mozilla for a real-world open design drive to improve Firefox accessibility. This work first appeared as: Salehi, Niloufar, and Bernstein, Michael S. "Hive: Collective Design Through Network Rotation." *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW (2018): 151.

N. Salehi (✉)
School of Information, University of California, Berkeley,
313 South Hall # 4600, Berkeley, CA 94720-4600, USA
e-mail: salehi@ischool.berkeley.edu

M. S. Bernstein
Stanford University, Gates Computer Science 384, Stanford, CA 94305-9035, USA

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_6

1 Introduction

People with shared challenges often gather online to discuss and envision shared outcomes. Examples include grassroots Twitter movements to advocate for new harassment policies, Mechanical Turk workers advocating for fairer treatment (Salehi et al. 2015), and competitions to address global grand challenges (Introne et al. 2011; Lifshitz-Assaf. 2014). For a collective to make progress, its members must deliberate to understand and articulate the problem, then collectively imagine goals and paths forward to alternative futures. How might social systems support large scale collective design (Fig. 1)?

Open design efforts have long sought to bring together collectives online to ideate, but in practice, participants rarely work together to understand others' viewpoints or build constructively on each others' ideas. Recent work has aimed to increase collaboration by intermixing ideas, for example by grouping and combining submissions (Siangliulue et al. 2015a, 2016; Yu and Nickerson 2011), by encouraging contributors to use each others' submissions (Boudreau and Lakhani 2015; Malone et al. 2017), or by exposing contributors to each others' high-level approaches (Yu et al. 2014a). This work all shares an underlying assumption that the information—the idea itself—is central. However, knowledge is situated and developed in the context of social relations, and organizational behavior research has demonstrated that mere exposure to others' ideas does not result in meaningful engagement with those ideas (Choi and Thompson 2005; Rink 2013; Rogers 2010). The result is open design efforts with far less intermixing and collaboration than originally envisioned (Bjelland and Wood 2008).

In this paper, we suggest an alternative approach: rather than intermixing ideas, intermix the people who bring the ideas. Research shows that collaboration with people who bring new perspectives results in engagement and increases the influence of new ideas, in a way that abstract exposure to those ideas does not (Choi and Thompson 2005; Rink 2013). A new person's addition to a group forces existing group members to re-evaluate hidden assumptions (Lin and Schwartz 2003) and overcome fixation (Choi and Thompson 2005) by engaging with the new member. Higher tie strength, trust, and acceptance between actors increase that influence (Rogers 2010).

Our goal with this research is to introduce a new class of social computing system that iteratively weaves a collaboration network by bringing members into contact with

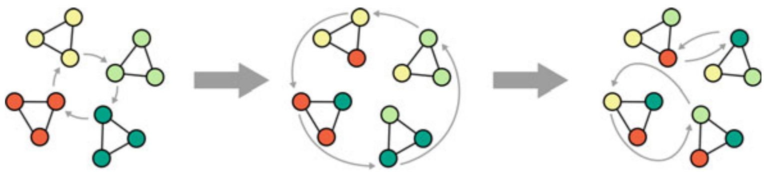


Fig. 1 Hive facilitates engagement with diverse viewpoints by rotating team membership in a collective over time. We introduce algorithmically-mediated *network rotation* to manage who should move, and when, to bring positive external influence to a team

diverse perspectives over time. We present *Hive*, a system that organizes a collective into teams and rotates people in the collective across teams at regular intervals as they work towards a shared goal. Hive connects people together from distant parts of the collaboration network to facilitate the spread of viewpoints within the collective. Underlying Hive is a new approach that we term *network rotation*, which changes team membership while maintaining high levels of tie strength within teams. To support network rotation, Hive models a collaboration network as a weighted graph where each edge captures the tie strength between two people. We use this collaboration network to identify who should be moved onto other teams and when.

Intuitively, we prefer a tightly knit network where most people are connected through a small number of links to a disconnected network with silos and bridges. To achieve a good network structure, we want to make a small number of moves that tighten up the collaboration network by bringing together people who are far away from each other. We frame this as an optimization problem where we maximize a weighted sum of two functions. First is *network efficiency*, or the average path length between two nodes in the collaboration graph, which captures how efficiently viewpoints and ideas can spread in the network. Second is *tie strength*, or the average edge weight between people who are in the same team, which captures how familiar team members are. The optimization framework considers them jointly: moving people between teams increases network efficiency but reduces tie strength, so the algorithm must trade these off. This is a large combinatorial search space, and cannot be explored through traditional means. We introduce a stochastic search algorithm that finds an effective solution by exploring thousands of possible moves.

Our evaluation asked whether Hive and network rotation led to effective design, both in a field experiment and in a public deployment with accessibility advocates at Mozilla. For the field experiment, we recruited 115 participants from Amazon Mechanical Turk to work together over three days to design proposals for an open-ended design challenge: How might we design a neighborhood common space to be a place that brings people together to foster strong community and mutual aid? We randomly assigned participants to one of three conditions: control, network rotation, and efficiency. A judge who was an expert in design thinking compared the concepts submitted in each condition, blind to condition. Proposals submitted by teams in the network rotation condition were rated significantly higher than control teams, whereas teams in the efficiency condition were not statistically distinguishable from those by control teams. Analysis of the chat logs suggests that rotation did shift the topic of teams' discussion, supporting the proposed mechanism. Qualitative and survey results indicated that membership change caused helpful disruptions without measurably decreasing members' psychological safety or interrupting social dynamics.

We then sought to understand whether the results from the field experiment generalized to a real-world deployment: Mozilla used Hive to run an open design drive on Firefox accessibility. Over one hundred volunteer disability experts, designers, and programmers collaborated on a week-long design drive, resulting in 60 proposals. Mozilla supported the proposals and plans to move ahead with prototyping and deploying five of them. Participants also enjoyed the diversity of viewpoints and fresh perspectives brought in through membership change.

Our main contribution in this paper is a system for supporting collective design by computationally adapting the collaboration network over time. We contribute network rotation, an algorithm that creates the conditions for people to engage deeply with a small group while benefiting from the scale and diversity of the collective and building social ties across the network. Hive is a step toward social systems that model and center peoples' relationships to one another.

2 Related Work

Hive presents a model of collective design that fuses the dynamics of close teamwork with the scale of online collectives, with the goal of engaging a diverse range of viewpoints through membership change. We draw on social psychology and organizational behavior research studying the conditions for open-ended, creative collaboration. We base our approach to collective design on traditions of human centered and participatory design.

Hive is a system that coordinates team members to achieve a shared goal by tracking and shaping the underlying social network. Prior work has designed algorithms for online team formation in social networks that focus on team composition, for example organizing teams based on skills required for the task, communication overhead, workload, and compatibility (Anagnostopoulos et al. 2012; Lappas 2009; Lykourentzou et al. 2017). Hive builds on this line of research by focusing not on the initial composition of the team, but on the dynamics of membership: how to balance both local team effectiveness and global network structure.

2.1 Membership Change Brings New Insight to Teams

Brainstorming, the practice of creative idea generation, relies on the cognitive stimulation facilitated by multiple members of a group, rather than individuals in isolation (Paulus and Yang 2000; Sutton and Hargadon 1996). Interacting with others can spark new ideas that individuals may not achieve alone (Brown 1998). By working together in a team, members build common ground, learn to coordinate, and discover how to utilize each person's unique skills (Liang et al. 1995; Reagans et al. 2005).

However, without changes in composition, teams risk becoming complacent, isolated, and less innovative (Bantel and Jackson 1989; Jackson 1996). New members bring novel perspectives to stale teams that can stir creativity. In fact, major innovations most often occur when new members join a team (Ettlie 1980) and innovation can come from both newcomers and novel combinations of old-timers (Perretti and Negro 2007). Even when the new member's conflicting points and opinions are flawed, the act of raising these points nevertheless leads to new insights in groups (Nemeth 1992). Our model of network rotation relies on this intuition.

There are costs to altering team membership. First, the process of adding a new team member can be disruptive to the working processes of a team, especially at high turnover rates (Cini 2001; Price 1977). If team membership is algorithmically determined, the team will wonder why the system made the decisions it did (Jahanbakhsh et al. 2017). Additionally, familiar teams develop psychological safety over time (Gruenfeld et al. 1996; Siemsen 2009). Disrupting this stability could have harmful implications for the team, because psychological safety fosters a team environment where members are comfortable taking risks with each other (Edmondson 2002). Without psychological safety, strong norms around conformity discourage new members from offering alternative perspectives before they are fully assimilated, leading to missed opportunities for generating and obtaining creative insight from an external source (Nemeth and Staw 1989; Gadon 1988). New members too are less likely to impose their ideas and to initially observe the group, with the goals of understanding the existing group dynamics and gaining eventual acceptance by the group (Heiss and Nash 1967; Nash and Wolfe 1957; Feldbaum et al. 1980).

Hive amplifies the concepts behind this literature, introducing the first system to facilitate membership rotation. It aims to support the team processes that this literature has found to be successful. Hive facilitates membership rotation with a global perspective of the network, enabling it to bring new perspectives in contact with each other. We draw on the insights from this literature to design for the interplay between the group's dynamics and the new member's fresh perspectives, avoiding rotating members too quickly and enabling team members to build up familiarity before moving.

2.2 Open Innovation and Crowdsourced Design

Large-scale innovation platforms and initiatives—such as IBM's "Innovation Jam" for product ideas, the U.S. government's change.gov site, and Google's 10 to 100th project for charity—promise to gather people to find innovative solutions to challenging problems. This solution has been used by large organizations looking for external sources of innovative ideas. However, the majority of the ideas that have been generated are deemed repetitive, commonplace, and unwieldy in volume (Bjelland and Wood 2008; Klein and Garcia 2015; Riedl et al. 2010; Siangliulue et al. 2016)—not a collaboration among a diversity of viewpoints.

The cost of sifting through such high volumes of low quality ideas often does not justify the benefit: IBM had 100 senior executives spend several weeks filtering ideas, the government shut down change.gov prior to its intended end date due to participation levels that overwhelmed the system, and 3000 Google employees were put 9 months behind schedule by managing the enormous mass of submissions (Bjelland and Wood 2008; Klein and Garcia 2015). To address this, platforms have introduced systems to more efficiently surface high quality submissions; nevertheless, filtering accuracy is still an issue and often the best results are those that are non-obvious or may not be rated most highly (Klein and Garcia 2015). Instead of extracting large volumes of individual ideas and evaluating them post-hoc, we propose creating the conditions for participants to engage with each other and collaborate.

One method for fostering collaboration in generating new ideas is by exposing people to each others' ideas. Crowd-driven idea maps (Siangliulue et al. 2015b) and analogies (Yu et al. 2014a, 2016a) connect contributors to concepts that they haven't seen before, prompting new ideas (Chan et al. 2016a). Enabling a traversal through the space of ideas requires extra metadata, but crowds can author this metadata as they work (Siangliulue et al. 2016) or it can be learned passively by aggregating behavior traces (Talton et al 2009). Once this metadata has been authored, people search through solution spaces at a higher level of abstraction (Yu et al. 2014b; Siangliulue et al. 2015b) and embed each other's solutions directly into their own (Malone et al. 2017; Monroy-Hernández and Resnick 2008; Nickerson 2015). Rather than putting people in contact with external *ideas*, our work presents an approach of putting people in touch with other *people*. Research has shown that people are more receptive to new ideas when they interact closely with those who bring the ideas (Choi and Thompson 2005).

Prior research has explored involving external crowds in design work. Crowds can provide feedback on in-progress designs (Xu et al 2015; Xu 2014; Luther et al. 2015; Mahyar et al. 2018), helping designers iterate more quickly. The crowd's feedback can be used to guide the crowd's next phase of ideation (Kim et al. 2017). However, the crowd is not always an expert critic (Wauck et al 2017): external expert facilitators can be effective at guiding the crowd in their ideation process (Chan et al. 2016b), or in setting a creative direction (Kim et al. 2014; Kulkarni et al. 2014). Crowds are also able to identify domains whose experts might be helpful for the problem at hand (Yu et al. 2016b). Hive expands on this point, combining it with knowledge that the best insights come from distant perspectives, not proximal ones (Wauck et al 2017), to create an algorithm that rotates distant members onto the team. Relative to this prior work, Hive does not view the crowd as static or external: we contribute a method that uses other members of the collective to provide outside perspectives, and continuously evolves the collaboration network to engage with those perspectives.

Taken together, this prior literature on open innovation and crowdsourced design systems has placed abstract ideas front and center—showcasing them, analyzing them, and remixing them. Hive contributes an alternative perspective of placing relationships front and center, and trusting good ideas to arise from successful collaborative relationships.

3 Hive

Hive forms and evolves teams over time within a large collective to achieve an open-ended goal such as collective design. In this section, we describe a scenario of how Hive can be used to support collective design. Then, we will detail the network rotation algorithm that underpins Hive.

Suppose that a large collective is interested in a design goal, for example improving the web browsing experience to better account for disability and inclusion. All participants sign up on Hive. Hive invites them to Slack and keeps record of their Slack username so that it can communicate with them via the Slack API. An administrator sets up the Hive project, determining for example the team size (e.g., 3–5 people), the number of phases, and instructions to give teams at each phase. In this scenario, the administrator creates phases for each stage of the design thinking process based on the OpenIDEO process: inspire, define, ideate, prototype, and test (Plattner 2010).

Hive enables identification of team leads, who act as facilitators. Facilitators are central to successful and satisfying online collaborations (Garrison 2006; Oxley et al. 1996). The team lead's role is to thread discussions, schedule meeting times, onboard new members, sustain commitment, capture ideas, and make sure that the daily deliverables are submitted. To maintain organizational memory, Hive does not rotate team leads across teams.

Hive then splits the large collective into many small teams with one team lead each. Hive notifies each person on Slack of their team, creates a channel on Slack for the team, and invites members to the channel. Each team is assigned a Google Drive folder where the instructions for each phase are in a shared Google document. Hive notifies the team of the instructions for the first phase by sharing a link to the folder with them. The instructions provide resources and specify the process and deliverables. The teams then work collaboratively for the duration of the phase. They collaborate using shared document editors, text chat, and photos of sketches.

Hive is now ready to begin the network rotation process and start the next phase. Hive's network rotation algorithm identifies who to move and where to move them, based on the collaboration network formed by the team groupings in all rounds so far. The administrator triggers Hive to execute the rotation, and Hive sends a message to each rotated member's old team channel and to their new team's channel encouraging a welcome for the new team member. Then, it removes rotating members from their original team channels and adds them to their new teams' channels (Fig. 2).

The process continues, with Hive rotating team members in each new phase to diffuse influence and ideas around the network. Next we describe Hive's network rotation algorithm.

4 Network Rotation

Hive's central goal is to create the conditions for people to engage with a diversity of new perspectives while maintaining a level of within-team tie strength that can foster meaningful collaboration. To do so, Hive gradually changes team membership through a process we call *network rotation* (Fig. 4). Network rotation requires an algorithm that computes who to move, and where to move them—an exponential search space. In this section we first describe how Hive models the collaboration

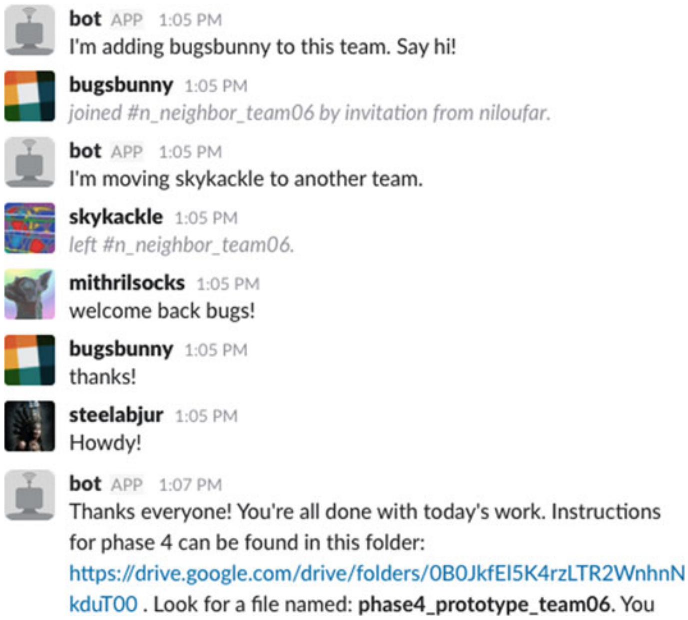


Fig. 2 Hive's Slack bot performs membership changes

network, then we describe an optimization objective that allows us to measure the effectiveness of a given rotation, and finally we describe a stochastic search algorithm that we use to execute that optimization.

4.1 Constructing a Collaboration Network

To support network rotation, Hive requires a model of the collaboration network as it evolves over time. The collaboration network is encoded as a weighted graph where the weight of each edge between i and j captures the tie strength between person i and j. We define tie strength in the range [0, 1], with 0 indicating a relationship between two strangers and 1 indicating an extremely strong tie, with other relationships falling somewhere in between (Gilbert and Karahalios 2009).

Our model must capture how new ties change over time. Network rotation can operate with any model of tie strength evolution over time. We offer one model here as a proof of concept based on prior literature. Tie strength and how it changes over time relies on many different social factors that are difficult to isolate (Marsden and Campbell 1984). However, prior literature generally assumes an upper limit, a decrease in tie strength due to lack of communication, and a positive relationship with time spent (Marsden and Campbell 1984; Burke and Kraut 2014). Other factors also play a role such as the emotional intensity of a relationship, however due to the

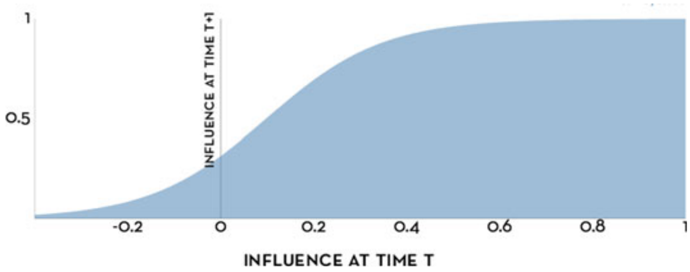


Fig. 3 We apply a logistic function that increases the tie strength between two people who work together in a team. We choose a logistic function as a reasonable proxy for tie strength because it captures relationships that are accelerate rapidly at first then asymptote, similar to familiarity in the real world (Katz and Allen 1982). For example, if two teammates have never worked together and have tie-strength of 0, one day of working together will increase it to $logit(0) = 0.16$

limited time spent in our studies and for the sake of simplicity, we rely solely on time spent communicating within our system.

For each i and j who were on a team together at time t, we apply a function that increases the edge weight between i and j. To model the growth of tie strength over time, we use a logistic function (Fig. 3). Logistic functions are commonly used to model natural and social processes that grow exponentially at first but that reach a saturation point and stop growing, such as diffusion of innovations (Marsden and Campbell 1984). A logistic function has also been used to predict the strength of persistent ties over time (Navarro et al. 2017). Likewise, with collaboration, teammates begin weakly connected, achieve increased performance through familiarity (Salehi et al. 2017), but eventually stop improving (Katz and Allen 1982). The constants of the logistic function can be tuned to impact how long teammates work together before their tie strength is high.

Second, the model must capture how the strength of past ties dampens over time. We update the collaboration network at every time step t, based on the collaboration network at time $t - 1$. We fade tie strength from prior time steps by first multiplying all edge weights by a constant dampening factor λ: $0 < \lambda < 1$. This multiplier retains old familiarity, but weakens it.

In our deployments, we used a logistic function with $k = 8$ and a dampening factor of $\lambda = .8$ to achieve a reasonable model of tie strength increasing each day over a week. For deployments with more phases, lowering k and increasing λ would slow down these dynamics.

4.2 Optimization: Tie Strength and Network Efficiency

Now, we use the collaboration graph to identify who should be moved onto other teams to bring new perspectives. Intuitively, we want to make moves that tighten

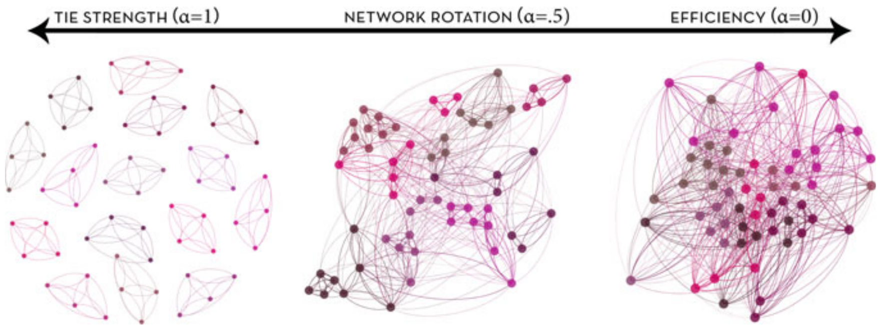


Fig. 4 Network rotation must balance network efficiency against tie strength in the collaboration graph. Left: Caring only about tie strength prompts the algorithm to create static teams. Right: Caring only about network efficiency breaks apart teams each round, which is highly disruptive. Middle: our algorithm balances the two goals, maximizing network efficiency while minimizing disruption

up the collaboration graph by bringing together people who are far away from each other in the collaboration network, but to avoid making so many moves that it is disruptive to the teams. We frame this as an optimization problem where we maximize a weighted sum of two functions (Fig. 4). First is *network efficiency*, which captures how different viewpoints have intersected with each other so far. Network efficiency, or $\text{Efficiency}(G)$ for a collaboration graph G, is calculated as the average path length between every two nodes (Latora and Marchiori 2001). Second is average *tie strength*, which captures how familiar team members are. Average tie strength, or $\text{TieStrength}(G)$, is calculated as the average edge weight between nodes who are in the same team. Moving people between teams increases network efficiency but reduces tie strength, so the algorithm must trade these off. First, we normalize $\text{TieStrength}(G)$ by a constant value based on network size (in our deployment $c = 0.005$) so that it can be traded off against $\text{Efficiency}(G)$. Next, we formalize this tradeoff with a parameter α. This produces an optimization function f on the collaboration network G:

$$f(G) = \alpha * \text{TieStrength}(G) + (1 - \alpha) * \text{Efficiency}(G)$$

In our deployments, we set $\alpha = 0.5$, equally weighting both functions. Finding the right subset of all possible moves to maximize this function is the topic of the next section.

4.3 Stochastic Search

Finding an effective rotation represents an extremely complex search space—$\mathcal{O}(2^N)$—where we cannot afford a brute force search. Unlike most exponential search problems that have relatively tractable solutions with integer programming or related techniques, network rotation is particularly challenging because its objective is non-linear: we must update edge weights after each considered set of moves, which may shift prior network efficiency measures. The main source of complication in this problem is that it is impossible to know the value of a single move independently of all of the other moves that will happen. The value of moving a person to a new team depends on who else moves from or to that team, and to or from other teams across the network. For example, if person p moves from team t_1 to t_2, an algorithm can evaluate this single move's impact on tie strength and network efficiency. However, the algorithm may now wish to move someone from t_2 to another team as part of the rotation. When it does so, it may move the very person that made t_2 so attractive a location for p to join in the first place, removing the basis for putting p on t_2. So, if this were an integer program, it would require $\mathcal{O}(2^N)$ variables and have poor structure—not totally unimodular—resulting in modern solvers unable to solve it effectively.

In this section, we introduce a method for finding effective network rotations: a stochastic search algorithm that explores thousands of possible moves and chooses a set of moves that increase the value of $f(G)$. Stochastic search is a good fit for problem spaces that do not have a helpful structure affording traditional optimization techniques. Instead, stochastic search explores alternatives and follows a gradient with random restart. We propose this approach not as the only possible algorithmic solution, but to establish the fruitfulness of the approach.

Hive's stochastic search algorithm calculates a network rotation by finding a set of moves S that create a new graph G maximizing the function $f(G)$. To do so, it considers a set of all possible moves, picks one probabilisitically, evaluates the optimization function after that move, and adds it to the current solution if it increases the optimization function $f(G)$. The algorithm repeats these steps with a random restart to ensure that it explores many possible alternatives.

Algorithm 1 describes the stochastic search algorithm. Take $S_{current}$ to be the best solution so far. At the start, $S_{current}$ is an empty set. The algorithm adds a valid move to the set at each step. The function Transform(G, S) returns a graph G' that is the result of performing the moves in the set S on the graph G. The algorithm can then compare $f(G')$ with the current best solution and update it. To prevent getting stuck in local maxima, with a probability of $\epsilon = 2 \times 10^{-6}$ at each step, the algorithm does a random restart. Hive runs this algorithm repeatedly and takes the best solution out of all of these iterations. In our deployment we set the algorithm to run one thousand times for a reasonable run time.

In order to add a valid move to the set S, the algorithm first creates a list of all moves of one person to another team. We constrain this set to valid moves, which cannot involve a team lead and cannot put a team's size outside the administrator's limits (Hive's default is 3–5 people). We include an additional optimization: the algorithm might move person p to team t_1, and then move that same person to another team t_2, which would be the same as moving p once to t_2. To reduce the search space, valid moves can only move each person once per round. Finally, the algorithm picks a move at random from the set of valid moves, creating a new graph G', and evaluates the optimization function $f(G')$ on that graph.

Algorithm 1 Network Rotation with Stochastic Search

procedure STOCHASTICSEARCH(G)
  $S_{current} \leftarrow \{\}$
  while true **do**
    $S_{candidate} \leftarrow \text{AddValidMove}(S_{current})$
    $G' \leftarrow \text{Transform}(G, S_{candidate})$
    $G_{current} \leftarrow \text{Transform}(G, S_{current})$
    if $f(G') > f(G_{current})$ **then**
      $S_{current} = S_{candidate}$
    if Random() $\leq \epsilon$ **then**
      return $S_{current}$

5 Field Experiment

Our goal in evaluating Hive is to answer the following questions: (i) Is network rotation more effective at supporting collective design than no rotation (i.e. control) or rotation for efficiency? (ii) Are large online collectives able to use Hive to follow the design process and collaborate on a real-world design goal? (iii) What behavior patterns emerge in small, rotating design teams of this nature? To answer these questions, we first ran a controlled study comparing Hive's network rotation algorithm against two other conditions: control (static teams), and network efficiency (rotation not constrained by tie strength). Then, Mozilla used Hive to host an online design drive on Firefox accessibility. We report on this field deployment to demonstrate how Hive could work in a real-world setting.

5.1 Method

We recruited participants on Amazon Mechanical Turk (AMT). To prevent language and technology use barriers, we limited our pool to people from the U.S. or Canada who have completed more than 500 tasks on AMT and have above a 95% approval rate. This process produced $N = 115$ participants. The gender balance was 55% female, 44% male, and 1% identified as neither or did not respond. Ages: 41% were 35–44, 40% were 25–34, 8% were 45–54, 5% were 18–24, and 3% were 55–64 years old.

We gave participants the following design prompt:

> How can we design a neighborhood common space to be a place that brings people together to foster community and mutual aid? If your neighborhood was given a space roughly the size of an elementary school playground, what could you do with it? Our goal is to create proposals for spaces/activities that bring together neighbors and strengthen communities.

We chose this prompt because most people have direct experiences with the topic of neighborhoods and shared spaces.

We paid participants an average of $99.40 for their participation of around 6 hours ($SD = 7.4$). The levels of commitment and the time participants spent on the project varied across teams. We used a combination of peer-reviews and evaluation of final deliverables to ensure fair payment: In line with the Dynamo guidelines,[1] we initially guaranteed a payment of $90 to anyone who would complete the project and a $50 bonus to team leads, who were responsible for creating storyboards/prototypes. Participants could also distribute credit to others who they felt deserved it, which we used along with expert evaluation of final proposals to pay bonuses.

In designing our field experiment we aimed to study three different optimization objectives for Hive's network rotation algorithm. First is tie strength ($\alpha = 1$), which maximizes tie strength within teams. This approach keeps all teams fixed. Second is network efficiency ($\alpha = 0$), which maximizes network efficiency by making moves that diffuse team members around the network, but disregards tie strength. Finally, balanced ($\alpha = 0.5$), which jointly maximizes tie strength and network efficiency (Fig. 4).

We randomly assigned participants to one of three experimental conditions (Fig. 4). *Control* uses the tie strength algorithm to keep teams static. *Network efficiency* uses the network efficiency algorithm with a limit on the number of moves. *Network rotation* uses the network rotation algorithm to balance the other two conditions. The major difference between the latter two conditions is that network efficiency is not trying to achieve high tie strength within teams.

In simulations on networks the size of our study, network efficiency ($\alpha = 0$) made a large number of moves, 52 moves on average. Our pilots and prior research demonstrated that this level of membership change was very disruptive to teams (Salehi et al. 2017) and failed to have compelling results, making it too weak a control condition. Therefore we decided to limit the number of moves. The network rotation condition made around 27 moves, which was more feasible for the teams to manage, so we limited network efficiency to 27 moves in order to compare the effect of the algorithm objective when given the same budget of number of moves.

When participants signed up for the study, we asked them to perform an initial brainstorm and submit three ideas. We asked an independent reviewer with design experience to evaluate the ideas on a 1–5 rubric. We averaged the three scores and later used this baseline creativity score to control for baseline team member creativity in our analysis.

Hive organized members into teams of four within their condition, resulting in nine teams in control,[2] ten in network rotation, and ten in network efficiency. Hive then created a Slack channel for each team and guided them through a four-phase human-centered design process—empathize, define, ideate, prototype—spending one hour on each of the first three phases and three hours on the final phase. We scheduled the time for the first three phases based on participants' availability that they had provided when they signed up. Given the length of time required for the final prototype phase, teams were asked to make their own scheduling and collaboration decisions. Each

[1] http://guidelines.wearedynamo.org.

[2] One team lead dropped out of the study at the outset, so the control condition had one fewer team.

team had a team lead, who did not participate in membership changes. We provided teams with tutorials, educational videos, examples, and steps to take for each phase.

After each of the empathize, define, and ideate phases, Hive moved participants or kept them static based on the group's condition. After every phase, we asked participants to fill out a survey containing questions sourced from the psychological safety index (Edmondson 2013, 1999), rating a 1–7 Likert scale for items such as: "Members of this team are able to bring up problems and tough issues" and "Working with members of this team, my unique skills and talents are valued and utilized." Our goal was to measure how much psychological safety—a measure of perceived ability to take risks in a team and a strong correlate of team performance—suffered when Hive changed team membership. Summing the responses on the seven-question index produced a psychological safety score for their team. We also asked participants to rate how well their team collaborated in each phase on a scale of 1 to 5, and we asked people who had experienced team changes to rate how well they (or the new member) integrated into the team.

We gathered 87 final design submissions from Hive teams, three submissions per team. We asked an independent reviewer with design experience to evaluate the ideas. We provided the reviewer with the evaluation criteria that we had given the teams. The rubric was based on one used by the OpenIDEO collective innovation platform, and included: understanding needs, choosing an important problem, finding a unique solution, demonstrating empathy, and proposing a feasible solution. To provide a more sensitive instrument then Likert-scale ratings, we utilized pairwise comparisons on allourideas.org. allourideas aggregated the pairwise rankings to assign a composite score per project.

For all analyses, submissions were anonymized and raters were blind to condition. The rating method followed a method of alignment on a training set and then independent coding (e.g., (Salehi et al. 2017; Bernstein et al. 2012)). An author who had not seen the teams' projects and the external rater independently reviewed ten proposals blind to team and blind to condition, then clarified the rubric based on discussion. The external rater then proceeded to rate the full dataset.

We also measured the mechanism through which the effect was predicted to occur: specifically, whether the addition of the new member changed what the teams discussed, dislodging the team from their existing design trajectory or fixation. Prior theory predicts that teams fully integrating new members would change their ideas, rather than continue on their existing path (Choi and Thompson 2005). This shift should be visible in what the team discussed. We built a bag-of-words vector space model of the language used by each team in each phase. Text was pre-processed to remove stop words, lemmatized, and stemmed; we then computed the vector sum of the words as a representation of the team's language in the phase. If teams continue talking about the same ideas, cosine similarity between phases would remain high; if teams shift, cosine similarity would lower.

For qualitative analysis, we consolidated survey responses, deliverables, and chat logs according to the following pre-determined themes framed by our research question: following a collaborative design process, team dynamics, and the effect of network rotation.

5.2 Results

One hundred and fifteen participants completed the project in 29 teams and submitted 87 final proposals. Network rotation moved members around to diffuse their influence over the phases of the experiment (Fig. 5). Cumulatively, participants exchanged 125,000 messages on Slack. In this section, we first present a comparison of the teams' final ideas in each of the three conditions, and a test of the expected mechanism through which the difference would occur. We also detail the teams' processes and dynamics. The Discussion section will synthesize this study and the next to reflect on the conditions under which network rotation will succeed.

5.2.1 Network Rotation Resulted in High-Quality Proposals.

What effect did network rotation have on the proposals at the top of the score distribution? *The network rotation condition dominated the top-rated proposals.* The distribution of submissions can be seen in Fig. 6. In the top third of proposals as rated by the expert, half (17) were from network rotation, 9 were from network efficiency, and 6 were from control. A Chi Square test of independence comparing the number of proposals in each condition found a significant effect: $\chi^2(2) = 8.7$, $p < 0.05$. Inspecting the Pearson residuals confirmed that the network rotation condition contributed most heavily to the residuals in the model.

Considering the whole score distribution, the mean expert score in each condition, ranked descending, were network rotation ($M = 60$, $SD = 20$), network efficiency ($M = 47$, $SD = 23$), and control ($M = 46$, $SD = 17$). Inspecting the overall means, the network efficiency condition was nearly equivalent to the control

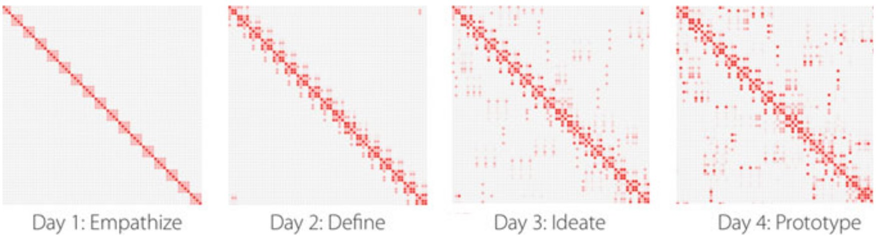


Fig. 5 The network rotation algorithm moved members of the collective into new positions to diffuse their influence. Darker cells in this adjacency matrix indicate higher tie strength

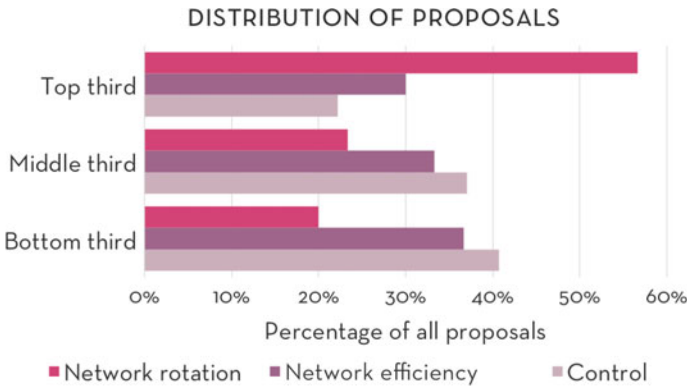


Fig. 6 Over half of the top-rated proposals were from the network rotation condition

condition (47 vs. 46), but network rotation was 0.7 standard deviations higher on average (60 vs. 46). We analyzed this data using a linear regression. We transformed the dependent variable using a box-cox transformation to ensure that the data met regression normality and homoskedasticity assumptions. We used the control condition, i.e. static teams, as the baseline intercept for the categorical condition variable, and expert-rated score as the dependent variable. We also controlled for each team's baseline creativity scores. We did this by calculating a weighted average of each team member's baseline creativity score, weighted by the number of rounds that member had been on the team. Replicating this analysis with untransformed data and no control variable leads to similar results. The regression confirmed the same results (Table 1): the network rotation condition was significantly better than the intercept control condition of static teams ($p < 0.05$), but network efficiency was not. A linear hypothesis comparison between the beta coefficients of network rotation and network efficiency is marginally significant ($p = 0.05$), strongly suggesting that network rotation teams outperformed network efficiency teams via this measure as well.

Qualitatively, surveys and chat logs suggest that the control teams suffered from design fixation, rarely looking past their initial brainstorm:

Table 1 Result of linear regression on score. ***: $p < 0.001$, *: $p < 0.05$. $N = 87$, Adj. R^2 is 0.04

Variable	β	SE	t
Network rotation	8.34*	3.86	2.16
Network efficiency	0.72	3.91	0.18
Baseline team creativity	1.62	2.38	0.68
Intercept	29.09***	8.46	3.44

> i didn't think this was a great team overall as our styles were too similar [...] i think new blood is sorely needed all around.

In contrast, network rotation caused teams to re-evaluate their syntheses. One participant, who convinced their new team to consider perspectives formed in their previous group, reported to us:

> I came up with a solution to a problem with one of the group's ideas. After that my contributions were taken into consideration [...] I came into a group with VERY different ideas than my original group.

Together, this quantitative and qualitative data suggest that network rotation introduced substantive, helpful changes to teams' ideation.

Membership change could subjectively feel costly for participants, and they sometimes complained to us about having to re-adjust and that they *"liked their previous team's ideas:"*. However, in most cases, teams were welcoming to the new member and they integrated quickly. When on-boarding worked well, the team appreciated the new ideas that the new member brought with them:

> - Thanks for your ideas [...] You had some of the best ever. It really gave us a head start.
> + I'm glad to have been helpful!! You guys had other awesome ideas too.

Membership change is most beneficial when the new member feels comfortable disagreeing with the rest of the team and proposing alternatives. However, newcomers may prefer to go along with the group in order to be accepted in the new team, resulting in groupthink that undermines the membership change. One way to prevent this is to decrease incentives for old-timers in the team to insist on established norms. For instance, when a new member is introduced as a "temporary" addition to a team, they are able to exert more influence than when they are introduced as a permanent addition (Rink and Ellemers 2009). In our experiment, we found that the assumption that teams would change over time made people more receptive to those changes and comfortable in raising conflicting ideas. For instance, in a pilot study on designing for disabilities one new member took experiences from their previous team and explained a new viewpoint to their new team:

> - [...] but I'm having a real though time trying to tell a disabled person how they feel when I've never been in their shoes.
> + I agree that we can't tell someone how they feel. We can only empathize with them.
> - I don't feel we're giving them enough credit. I don't think they want sympathy. I think they want a fair shot.
> * So that's a good place to work from...

New members often made an effort to understand the team's thought process. Since all deliverables were available in the shared Google Drive folder, new members sometimes went back to previous phases and read the team's work. They also read previous chat logs. Most frequently, new members asked questions and engaged in conversations with other team members. These conversations helped the team re-evaluate assumptions.

5.2.2 Language Changed More Within Groups Who Experienced Membership Changes.

Changes in the language are a signal of changing topics, and evidence that new membership shifted the topic of discussion. Teams engaged in an average of over 7000 lines of chat over the deployment.

In the rotation between the first two phases, teams in the network rotation and network efficiency conditions had lower cosine similarity scores in language compared to teams in the control condition (Fig. 7). A one-way ANOVA to compare the effect of experimental condition on the cosine similarity of text between the first and second phases had a significant effect of condition ($F(2, 26) = 3.4$, $p < .05$). In post-hoc planned contrasts comparing network rotation and network efficiency to control, the contrast with network rotation is significant ($p < .05$), and network efficiency is not ($n.s.$).

This result supports the proposed mechanism: that the addition of new members prompted integration of new perspectives and caused the team to shift ideas. Later phases showed no significant difference, suggesting that ideas may have begun to solidify by this point. However, the lack of difference in later phases despite membership change confirms that the language difference between the first two phases was not an artifact of teams welcoming new members.

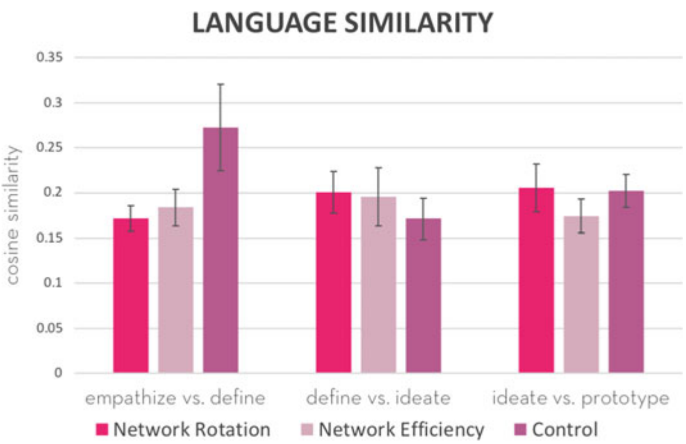


Fig. 7 Cosine similarity of the language used in teams' discussion between phases. Teams in the network rotation and network efficiency conditions shifted their discussion contents as a result of new membership, whereas teams with stable membership remained focused on the topics represented in their previous discussion

5.2.3 Psychological Safety Remained Mostly High as Teams Rotated

Despite the concerns that participants raised while rotating, psychological safety scores were relatively high and remained mostly at the same level throughout the project except for people who moved teams (Fig. 8). A two-way ANOVA with status (I moved/A team member moved/Static team) and phase as independent variables, and psychological safety score as dependent variable, had a significant effect of status ($F(2, 943) = 110$, $p < .001$) and a significant interaction effect ($F(4, 943) = 9.8$, $p < .001$). Inspecting individual cells, the significant interaction was driven by decreasing psychological safety scores for people who were moved on later days—indicating that it felt harder to join a team later on in the process. This result can be incorporated into the algorithm to make more changes early on or to have functionality for reversing changes that are too disruptive.

6 Field Deployment: Firefox Accessibility

Following the field experiment, we sought to better understand Hive in a real world setting. So, we collaborated with Mozilla as they used Hive for a week long open design drive with accessibility advocates. To complement the prior evaluation, this effort was a field study rather than a field experiment (Mcgrath 2000). Mozilla issued an open call via social media and over one hundred disability experts, designers, and programmers joined.

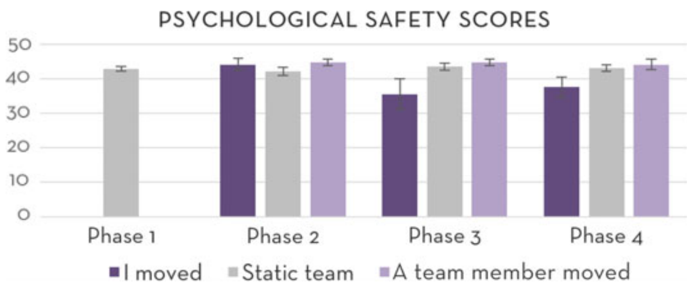


Fig. 8 Psychological safety remained high despite membership changes. In the overall study population we found no significant change in psychological safety scores. When we separated the different groups of participants we found that scores dropped for members who moved to a new team, but raised again for those members one round later, after the initial cost of reintegrating into the new team

6.1 Method

An open effort to design for accessibility needs strong voices from people who have experienced the challenges firsthand and understand points of friction and limitations. Therefore, we reached out to disability communities and to disabled individuals on AMT (Zyskowski et al. 2015), and invited them to join as team leads. Mozilla offered them a $150 gift card for lending their expertise to the project. We created a separate channel for team leads in which we shared information and instructions. All other participants volunteered their time.

Every participant filled out an initial form to apply to the project. We asked them for their basic demographic information, areas of expertise, and availability. Afterwards, we invited people to Slack and 113 people joined. Based on their time zone and availability, we divided the pool into five groups and placed people into teams within those groups.

Ninety-eight people completed the project in 23 teams. Of these, 42% had direct experience with disabilities and accessibility needs, 41% had design experience, and 13% had programming experience. Gender: 53% were female, 44% male, 2% preferred not to say, and 1% were gender queer. Location: 58 were from the U.S., 12 Canada, 8 India, 6 Australia, 2 Taiwan, 2 Sweden, 2 Germany, and 1 from Argentina, Brazil, France, Indonesia, Israel, the Netherlands, Nigeria, and the UK. Ages: 33% were 18–24, 26% were 25–34, 31% were 35–44, 7% were 45–54, and 3% were above 55. Over the weeklong project, 15 people (13%) dropped out due to scheduling conflicts or because they lost interest; the remaining 87% of participants completed the weeklong design drive.

The submitted designs were evaluated by an independent Mozilla accessibility panel. In addition, we sent out a final survey to participants asking for feedback on the process: "it is very important for us to hear about your experiences, the good and the bad." The survey consisted of the following open-ended questions:

- What did you like about the process?
- What can we do better in future design drives?
- What did you like about the membership changes?
- What do you wish we would do differently about the membership changes?

We consolidated survey responses, teams' daily deliverables, and chat logs according to the following pre-determined themes framed by our research question: participation in a real world collective design drive, team dynamics, and the effect of network rotation on teams.

6.2 Results

The effort resulted in over 60 proposals, which were evaluated by an independent Mozilla accessibility panel. In this section we first detail the process that the teams

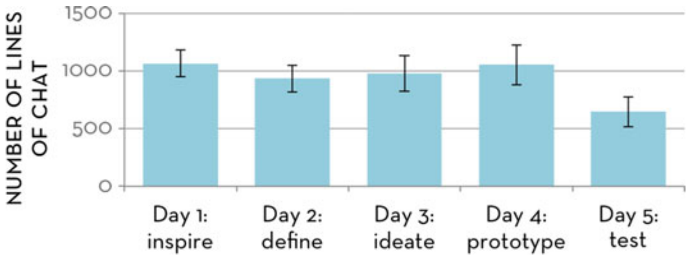


Fig. 9 Ninety-eight Mozilla volunteers completed the project in 23 teams. They exchanged 125,000 messages on slack over the course of 5 days. Testing required the least discussion in teams

went through, then we explain our qualitative findings from observing the process and analyzing survey responses (Fig. 9).

6.2.1 Design Stages

The phases of the design drive took place online over five days. *Day 1: Inspire.* Following OpenIDEO, the first phase "inspire" consisted mostly of storytelling and getting to know fellow team members. We asked participants to deliver a list of 10 obvious and 10 non-obvious observations that they had made in this phase. *Day 2: Define.* In this phase, we asked participants to identify patterns from the previous phase and define three different problem statements. *Day 3: Ideate.* In this phase, participants came up with 10 obvious and 10 non-obvious ideas to address their problem statements. We told them that it is likely that their problem statements would change as they ideate. Often a single submitted idea had been reshaped by multiple members who suggested insights and brought up constraints. Participants largely relied on their own experiences to ideate solutions that would be useful.

> Long day but I'm here and my grandparents just went to bed. Boy could my grand father use a lot of the ideas we came up with. I spent most of the night cleaning up his computer from all the pop ups that he's clicked on.

Day 4: Prototype. The prototyping phase was the most time-consuming and collaborative phase of the project. Participants chose three ideas from the previous phase. We suggested the guidelines: "the most likely to delight," "the rational choice," and "the most unexpected". They worked together to develop the idea using storyboards. Teams continued to evolve their ideas—often now with a new member—and discussed users and scenarios. Some team members drew rough sketches and photographed them to demonstrate what they meant. *Day 5: Test.* In the final phase, participants found a user and gathered feedback. They submitted observations from their testing, and updated ideas and storyboards to integrate new observations. Participants used creative ways to find users for testing. Some reached out to friends and family or support groups that they were a member of. Some participants who had access to disabled family and friends offered to help others by posting on the

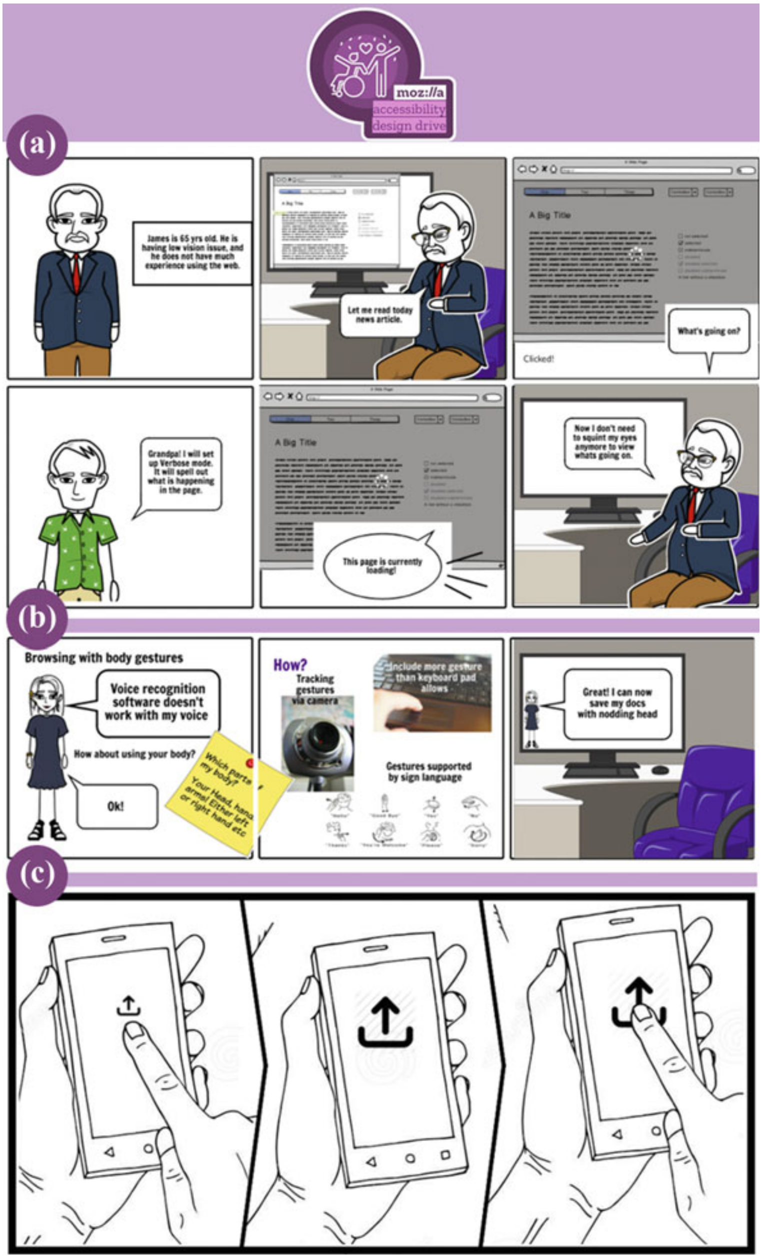


Fig. 10 Hive proposals submitted to the Mozilla accessibility design drive. **a** A browser feature that provides explanations for events without parallels to the non-digital world, e.g., loading times. **b** Deaf culture uses sign language call signs to adapt for more complex tasks. Here, a person using ASL can use their friend call sign to compose emails. **c** Browsers react when users are having trouble, for example, when a user tries to click on a button and misses, the button increases in size

collective's #general channel on Slack. This phase required more independent work and less collaboration than previous phases. Testing uncovered aspects of their ideas that teams had not considered before: for example, whether certain features would make browsing the web even more confusing, or what would happen when people with different needs and limitations share a computer. Participants added these observations and open questions to their proposals.

6.2.2 Mozilla and Participants Reacted to the Proposals Positively

While this second deployment was not focused on comparative evaluation against a control condition, it is still important to examine the quality of the final ideas. The independent Mozilla panel rated the final proposals positively, summarizing on their blog about the experience, "We were impressed with the results of this experiment," and that Mozilla would be pursuing five of the ideas for further testing and development.

We asked participants about their views on membership changes. Of the 35 participant responses to the final survey, 17 said that they liked the opportunity to engage with new perspectives, 8 said that they enjoyed getting to work with more people, 3 said they did not like it, and 3 had mixed feelings. Participants indicated that they felt network rotation helped facilitate strong ideas, brought "fresh perspectives" and "allowed for more ideas to come into the room":

> Meeting new people was great. Mixing ideas was very helpful, especially as people dropped in and out later in the week.

> I liked meeting others with different experiences and backgrounds. Everyone brought something interesting to the table.

> Wow, that didn't even feel like an hour! [...] I'm loving this! This is stuff I never talk about, it's helping me realize some things about myself that I was blind to before I started talking about it here. Thank you all.

However, it is also important to capture the challenges that participants faced while participating in network rotation. Participants fretted giving up teammates with whom they had developed stronger ties:

> I liked bringing in fresh ideas every day. Gave a different/new perspective [but] sometimes you build good working relationships with others and do not want to lose them.

Engaging with different backgrounds can be engaging and insightful, but also challenging (Lin and Schwartz 2003):

> I think there was a pretty big rift between the technical thinkers and the more arts and idea-based thinkers such as myself that we could never seem to bridge properly.

Often new team members went to the effort of reading past transcripts to understand the team's status, but this was not always the case:

> Sometimes it was difficult to get new members to read our previous discussions, so they didn't seem to understand where we were coming from.

The main functional request was for more control over membership changes: earlier notification, or the ability to prevent disruptive changes.

Overall, the feedback indicated that network rotation introduced positive stressors that were beneficial to the team and their ideas. Taken together, these results and observations reinforced the main conclusions of the field experiment, generalizing them to volunteers rather than paid Mechanical Turk workers.

7 Discussion and Future Work

Our primary objective in this paper was to create the conditions for collective design participants to understand others' viewpoints and build constructively on others' ideas by centering people and their relationships—rather than their abstract ideas. In doing so we contribute *network rotation* for organizing the collaboration through membership changes. In this section, we reflect on open questions and opportunities for our method. First, we consider the high level goal of network rotation: trading off local effectiveness with global desirability, and how those can be defined differently. Second, we analyze the conditions under which membership changes were effective. Finally, we describe the limitations of our current study and discuss design implications for collective design.

7.1 Trading Off Local Effectiveness with Global Desirability

Our main contribution in this paper is the concept of network rotation, which restructures the social network by balancing two needs: (1) local: an effective team, and (2) global: a desirable collaboration network. Hive's network rotation algorithm provides a lever to trade off these two measures. While we offer initial ideas of what such measurements might look like based on prior literature—logistic growth with time decay for tie strength over time, and network efficiency as a measure of the collaboration network—future work will improve on them. Further, any designer could change these measures to capture their goals for what an effective group looks like, and what a desirable collaboration network looks like, and network rotation would carry them out. Below, we analyze our choices and name a few alternatives concepts for these measures.

What constitutes a good team? In our evaluation we used a measure of team member familiarity (Salehi et al. 2017) by averaging pairwise tie strength of all people on the team. Other measures of tie-strength and how it changes over time can also be used. For the sake of simplicity and due to the relative short time frame of our study we based our measure of tie-strength on time spent collaborating. But not all time spent together is of equal significance. Researchers have included criteria such as emotion, communication intensity, overlapping social networks, shared identity, and the sporadic nature of communications within measures of tie strength (Marsden and

Campbell 1984; Navarro et al. 2017). Beyond tie strength, other factors for effective teams can include balancing personality types (Lykourentzou et al. 2016) or asking participants for feedback about their experiences with other team members (Salehi et al. 2017; Lykourentzou et al. 2017). Another is user input: does the team want to participate in rotation? (An open question: how much self-insight do teams have into when it would be most helpful to introduce new blood?). Linguistic signals such as quantity, qualities, and the temporal dynamics of communications between team members (Riedl and Woolley 2017) can also be used.

In addition to strong local collaborations, we care about the structure of the network globally. In our evaluation we aimed to encourage participants to engage with diverse viewpoints. To do so we initially randomly spread participants in the network and then aimed to maximize network efficiency, making the assumption that people with the longest path length between them were least likely to have heard each others viewpoints indirectly. This is an approximation of viewpoint diversity that we found worked well for the purpose of our evaluation (and provided a higher bar to show a significant effect). This was because participants did not join with preconceived designs for neighborhood common spaces or web accessibility opportunities. Therefore, even if participants were homogenous at first, they diverged quickly as soon as they started working together because each team came up with different ideas. We can imagine that in some other cases diversity of viewpoint may not be as easily proxied, for instance in cases of political, occupational, or socio-economic similarities among participants. In such cases, initial surveys can be used to prepopulate the network and spread people based on a stronger measure of viewpoint diversity. Beyond network efficiency, other factors may include structural diversity or the existence of internal boundary spanners in the group.

7.2 *Boundary Conditions of the Effectiveness of Network Rotation*

Given the results of our studies, what have we learned about when network rotation will produce positive results and when it might fail? We look to Tohidi et al. (2006) for a useful frame: they distinguish two phases of design: "getting the right design", early on when the idea is still malleable, and "getting the design right", later when the focus is refinement. Qualitative evidence from our deployments suggests network rotation was most effective early on, when getting the right design. Distant ideas caused teams to reflect critically on their assumptions (Nemeth 1986). Participants in the Mozilla deployment reported it let them "gain new perspectives on problems" and engage with "totally new ideas and working style[s]". Network rotation's value diminished later, when teams focused on getting the design right. Disruptions then became costly, and participants' suggestions involved restricting late moves ("Do it at the very beginning because it's hard to reintroduce again with the new team") and supporting continuity (sometimes a new member “didn’t seem to understand where we were coming from”).

In practice we found that the network rotation algorithm would sometimes move people back to a team that they had been a part of in the past, so that the algorithm could benefit from both the ties already established with the old members and the new gains in network efficiency. This is in line with prior literature that has found an increase in innovation both through the addition of newcomers and through novel combinations of old-timers (Perretti and Negro 2007). While our current network rotation algorithm does not differentiate between the two, future research can analyze the trade-offs between the addition of newcomers vs. “new” old-timers.

How small can a collective be before network rotation is not helpful? And how many phases are necessary to see a meaningful effect? When the collective is small, network rotation might make similar rotation selections as a random rotation, since the network will quickly become highly efficient and all moves will have the same efficiency value. Likewise, with too few rotations, the algorithm will not be able to weave the network together effectively. To investigate this, we performed simulations to test network rotation on varying group sizes and numbers of phases. Qualitatively, we observed that network rotation required a rough lower bound on group size (>50) and number of phases (>4): otherwise its changes to network structure were not meaningfully different than random rotation.

7.3 Limitations

One limitation of our study is that the deployment happened over a relatively short timescale of five days, which is far briefer than regular design efforts last. The result is that our study focused on early-stage concept development and left out iteration and evolution of projects for future work.

Our regression analysis confirmed that teams in the network rotation condition outperformed those in the control (static) condition, whereas those in the network efficiency condition did not. However, our study was limited in its ability to piece out the precise mechanisms of this effect—at what value of α does the effect begin to appear? A larger N study would allow us to perform a more detailed analysis with additional pairwise comparisons. Additionally, we rely on prior work to justify tie strength as an appropriate metric for local effectiveness of teams (Salehi et al. 2017), but we did not provide evidence. Psychological safety or other linguistic metrics may be used as signals to demonstrate the strength of local collaborations particularly when comparing the network rotation condition with network efficiency.

In the Mozilla deployment, teams were led by people who self-identified as disabled or were care-giver to someone with a disability. However, in practice we found that there were many more decisions about the process that were made by Mozilla, the leaders of the effort who had stakes involved: who can participate? When should someone be asked to leave? How will we share the results? Further research is needed to adapt the design process for collective online design and ensure that the communities involved benefit from the process.

Finally, a note for future work: due to the complexity of the search space for network rotation, all known methods require considering many or all subsets of moves at $\mathcal{O}(2^N)$, which is computationally intractable. Any solution requires an approximation, and our algorithm is one such approximation via monte carlo search. We do not claim optimality of this approach, and other approaches may exist that are equally or more effective. Future work will continue to examine this space.

7.4 Design Implications

Our deployments clarified avenues for improving the design of Hive and similar systems. Many analyses of human performance, from the Yerkes-Dodson Law (Yerkes and Dodson 1908) to the concept of flow (Csikszentmihalyi 1996), suggest that people perform best when facing a moderate amount of discomfort or challenge. Too easy and people zone out; too hard and they collapse under the pressure. Qualitative results made clear that network rotation introduced pressure and challenge to the design process: rotations forced people to get to know new teammates, question their previous assumptions and decisions, and potentially change their minds. However, our results also suggest that this discomfort led to better designs.

This puts Hive in an unusual position of being a design intervention that can actively create discomfort—future systems need to acknowledge this and design for it. User control remains an open question: teams wished for the ability to prevent changes. Future prototyping can answer the question of whether participants would use this ability to keep teams static when they would be better served by exposing themselves to new membership. One option would be to give each team one "token" to use throughout the process, allowing them to prevent changes once, when they deem it important, but set a norm of allowing rotation.

Given that teams were often reluctant to let a team member leave, there may exist alternative models that lessen the stress. One option would be for rotations to be sabbaticals rather than potentially permanent moves: the member would go join another team for one round, then return. Another would be to designate some members of the collective as rotation agents and not assign them to permanent teams, instead rotating them around between teams as temporary consultants. Both options would avoid breaking up the original team but still bring in new perspectives.

Moving forward, we hope to study the effects of large scale collaborative design teams over longer periods, for instance through multiple iterations of the design process. One question is what kind of long-term effects network rotation might produce. In addition, while moving people across teams, we observed that differences in inclusion varied across teams. Future work can explore how to onboard new members.

7.5 Ethical Implications

We argue that the ethical implications of this technology—and team management technologies more broadly—are directly related to the levels of autonomy and power that users have within the system and within the broader social and political context that the system is embedded in. In the best case scenario, these algorithms can provide effective means for collective self-organizing toward a shared goal. In the worst case they can alienate participants. For instance, rapid membership changes may be used deliberately to break social ties among workers and stifle collective action. We propose engaging in participatory design with potential users of these algorithmic systems as one strategy for social computing researchers moving forward.

8 Conclusion

In this paper we present a system that supports collective design. To do so, we argue that large collectives cannot cross-pollinate ideas simply by placing all ideas into the commons for others to see—people in the collective must instead interact with each other closely to benefit from ideas and perspectives that each bring to the table. We define an approach, network rotation, that supports this goal by gradually weaving together team membership from across the collective to increase the efficiency of the underlying collaboration network while maintaining high tie strength within teams. We embedded this approach in a system called Hive, and showed the value of network rotation in a controlled study. We deployed Hive with a Mozilla open design drive to demonstrate how our approach works in the real world.

When collective intelligence (Malone et al. 2009) and open innovation (Chesbrough 2006) gained popularity, the initial concept was to extract value from diverse, specialized innovators who work independently. The resulting successes span from Apple to Lego to NASA (Lakhani 2013). Our work challenges the assumption that collective design should occur through independent work. Instead, we suggest that design at scale should be *collaborative*. Our deployment demonstrated that people were often effective at dissemination, remixing, and engagement of ideas and viewpoints. Small teams have long been a vehicle for design thinking (Stempe and Badke-Schaub 2002). Drawing on recent work (Lykourentzou et al. 2016; Salehi et al. 2017), we propose that future systems continue to consider large-scale design as occurring in small-N teams within large-N crowds (Fig. 10).

Acknowledgements Special thanks to Turkers and Mozilla volunteers who participated in, and helped shape this project. This work was supported by a National Science Foundation award IIS-1351131, a Hasso Plattner Design Thinking Research Program grant, and a Stanford Graduate Fellowship.

References

Anagnostopoulos, A., et al. (2012). Online team formation in social networks. In *Proceedings of the 21st international conference on World Wide Web* (pp. 839–848). ACM.

Bantel, K. A., & Jackson, S. E. (1989). Top management and innovations in banking: Does the composition of the top team make a difference? *Strategic Management Journal*, *10*(S1), 107–124.

Bernstein, M. S., et al. (2012). Direct answers for search queries in the long tail. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 237–246). ACM.

Bjelland, O. M., & Wood, R. C. (2008). An inside view of IBM's 'Innovation Jam. *MIT Sloan Management Review*, *50*(1), 32.

Boudreau, K. J., & Lakhani, K. R. (2015). "Open" disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy*, *44*(1), 4–19.

Brown, V., et al. (1998). Modeling cognitive interactions during group brainstorming. *Small Group Research*, *29*(4), 495–526.

Burke, M., & Kraut, R. E. (2014). Growing closer on facebook: changes in tie strength through social network site use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 4187–4196). ACM.

Chan, J., Dang, S., & Dow, S. P. (2016a). "Comparing Different Sensemaking Approaches for Large-Scale Ideation". In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. CHI '16* (pp. 2717–2728). Santa Clara, California, USA: ACM. https://doi.org/10.1145/2858036.2858178. ISBN: 978-1-4503-3362-7.

Chan, J., Dang, S., & Dow, S. P. (2016b). IdeaGens: Enabling expert facilitation of crowd brainstorming. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion. CSCW '16 Companion* (pp. 13–16). San Francisco, California, USA: ACM. https://doi.org/10.1145/2818052.2874313. ISBN: 978-1-4503-3950-6.

Chesbrough, H. W. (2006). *Open innovation: The new imperative for creating and profiting from technology.* Harvard Business Press.

Choi, H.-S., & Thompson, L. (2005). Old wine in a new bottle: Impact of membership change on group creativity. *Organizational Behavior and Human Decision Processes*, *98*(2), 121–132.

Cini, M. A. (2001). Group newcomers: From disruption to innovation. *Group Facilitation*, *3*, 3.

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

Edmondson, A. C. (2002). Managing the risk of learning: Psychological safety in work teams. *International Handbook of Organizational Teamwork*, 1–38.

Edmondson, A. C. Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350. https://doi.org/10.2307/2666999. ISSN: 00018392.

Edmondson, A. C. (2013). *Teaming to innovate*. New York: Wiley.

Ettlie, J. E. (1980). Manpower ows and the innovation process. *Management Science*, *26*(11), 1086–1095.

Feldbaum, C. L., Christenson, T. E., & O'Neal, E. C. (1980). An observational study of the assimilation of the newcomer to the preschool. *Child Development*, 497–507.

Gadon, H. (1988). The newcomer and the ongoing work group. In *Team building: Blueprints for productivity and satisfaction* (pp. 161–175).

Garrison, D. R. (2006). Online collaboration principles. *Journal of Asynchronous Learning Networks*, *10*(1), 25–34.

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 211–220). ACM.

Gruenfeld, D. H., et al. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes*, *67*(1), 1–15.

Heiss, J., & Nash, D. (1967). The stranger in laboratory culture revisited. *Human Organization*, *26*(1–2), 47–51.

Introne, J., et al. (2011). The climate CoLab: Large scale model-based collaborative planning. In *2011 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 40–47). IEEE.

Jackson, S. E. (1996). The consequences of diversity in multidisciplinary work teams. In *Handbook of work group psychology* (pp. 53–75).

Jahanbakhsh, F., et al. (2017). You want me to work with who?: Stakeholder perceptions of automated team formation in project-based courses. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 3201–3212). ACM.

Katz, R., & Allen, T. J. (1982). Investigating the not invented here (NIH) syndrome: A look at the performance, tenure, and communication patterns of 50 R&D project groups. *R&D Management*, *12*(1), 7–20. https://doi.org/10.1111/j.1467-9310.1982.tb00478.x ISSN: 1467-9310.

Kim, J., Cheng, J., & Bernstein, M. S. (2014). Ensemble: Exploring complementary strengths of leaders and crowds in creative collaboration. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 745–755). ACM.

Kim, J., et al. (2017). Mechanical novel: Crowdsourcing complex work through reaction and revision. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. CSCW '17* (pp. 233–245). Portland, Oregon, USA: ACM. https://doi.org/10.1145/2998181.2998196. ISBN: 978-1-4503-4335-0.

Klein, M., & Garcia, A. C. B. (2015). High-speed idea filtering with the bag of lemons. *Decision Support Systems*, *78*, 39–50.

Kulkarni, A., et al. (2014). Wish: Amplifying creative ability with expert crowds. In *Second AAAI Conference on Human Computation and Crowdsourcing*.

Lakhani, K. R., Lifshitz-Assaf, H., & Tushman, M. (2013). Open innovation and organizational boundaries: Task decomposition, knowledge distribution and the locus of innovation. In *Handbook of economic organization: Integrating economic and organizational theory* (pp. 355– 382).

Lappas, T., Liu, K., & Terzi, E. (2009). Finding a team of experts in social networks. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 467–476). ACM.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, *87*(19), 198701.

Liang, D. W., Moreland, R., & Argote, L. (1995). Group versus individual training and group performance: The mediating role of transactive memory. *Personality and Social Psychology Bulletin*, *21*(4), 384–393.

Lifshitz-Assaf, H. (2014). From problem solvers to solution seekers: dismantling knowledge boundaries at NASA.

Lin, X., & Schwartz, D. L. (2003). Reaction at the crossroads of cultures. *Mind, Culture, and Activity*, *10*(1), 9–25.

Luther, K., et al. (2015). Structuring, aggregating, and evaluating crowdsourced design critique. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 473– 485). ACM.

Lykourentzou, I., et al. (2016). Personality matters: Balancing for personality types leads to better outcomes for crowd teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 260–273). ACM.

Lykourentzou, I., Kraut, R. E., & Dow, S. P. (2017). Team dating leads to better online ad hoc collaborations. In *CSCW* (pp. 2330–2343).

Mahyar, N., et al. (2018). CommunityCrit: Inviting the public to improve and evaluate urban design ideas through micro-activities.

Malone, T. W., et al. (2017). Putting the pieces back together again: Contest webs for large-scale problem solving.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2009). Harnessing crowds: Mapping the genome of collective intelligence.

Marsden, P. V., & Campbell, K. E. (1984). Measuring tie strength. *Social Forces*, *63*(2), 482–501.
Mcgrath, E. (2000). Methodology matters: Doing research in the behavioral and social sciences. In *Readings in human-computer interaction: Toward the year*, (2nd ed. 1995).
Monroy-Hernández, A., & Resnick, M. (2008). Empowering kids to create and share programmable media. *Interactions*, *15*(2), 50–53.
Nash, D. J., & Wolfe, A. W. (1957). The stranger in laboratory culture. *American Sociological Review*, *22*(4), 400–405.
Navarro, H., et al. (2017). Temporal patterns behind the strength of persistent ties. *EPJ Data Science*, *6*(1), 31.
Nemeth, C. J. (1986). Differential contributions of majority and minority inuence. *Psychological Review*, *93*(1), 23.
Nemeth, C. J. (1992). Minority dissent as a stimulant to group performance. *Group Process and Productivity*, *95*, 111.
Nemeth, C. J., & Staw, B. M. (1989). The tradeoffs of social control and innovation in groups and organizations. *Advances in Experimental Social Psychology*, *22*, 175–210.
Nickerson, J. V. (2015). Collective design: Remixing and visibility. In *Design computing and cognition* (pp. 263–276). Springer.
Oxley, N. L., Dzindolet, M. T., & Paulus, P. B. (1996). The effects of facilitators on the performance of brainstorming groups. *Journal of Social Behavior and Personality*, *11*(4), 633.
Paulus, P. B., & Yang, H.-C. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes*, *82*(1), 76–87. https://doi.org/10.1006/obhd.2000.2888. ISSN: 0749-5978.
Perretti, F., & Negro, G. (2007). Mixing genres and matching people: A study in innovation and team composition in Hollywood. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *28*(5), 563–586.
Plattner, H. (2010). *An introduction to design thinking process guide*. Stanford: In The institute of design at Stanford.
Price, J. L. (1977). *The study of turnover*. Iowa State Press.
Reagans, R., Argote, L., & Brooks, D. (2005). Individual experience and experience working together: Predicting learning rates from knowing who knows what and knowing how to work together. *Management Science*, *51*(6), 869–881.
Riedl, C., & Woolley, A. W. (2017). Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and emergent collaboration to crowd-based problem solving performance. *Academy of Management Discoveries, 3*(4), 382–403.
Riedl, C., et al. (2010). Rating scales for collective intelligence in innovation communities: Why quick and easy decision making does not get it right.
Rink, F. A., et al. (2013). Team receptivity to newcomers: Five decades of evidence and future research themes. *The Academy of Management Annals*, *7*(1), 247–293.
Rink, F. A., & Ellemers, N. (2009). Temporary versus permanent group membership: How the future prospects of newcomers affect newcomer acceptance and newcomer influence. *Personality and Social Psychology Bulletin*, *35*(6), 764–775.
Rogers, E. M. (2010). *Diffusion of innovations*. Simon and Schuster.
Salehi, N., et al. (2015). We are dynamo: Overcoming stalling and friction in collective action for crowd workers. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1621–1630). ACM.
Salehi, N., et al. Huddler: Convening stable and familiar crowd teams despite unpredictable availability. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. CSCW '17* (pp. 1700– 1713). Portland, Oregon, USA: ACM. https://doi.org/10.1145/2998181.2998300. ISBN: 978-1-4503-4335-0.
Siangliulue, P., et al. (2015a). Providing timely examples improves the quantity and quality of generated ideas. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (pp. 83–92). ACM.

Siangliulue, P., et al. (2015b). Toward collaborative ideation at scale: Leveraging ideas from others to generate more creative and diverse ideas. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 937–945). ACM.

Siangliulue, P., et al. (2016). IdeaHound: Improving large-scale collaborative ideation with crowd-powered real-time semantic modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 609–624). ACM.

Siemsen, E., et al. (2009). The influence of psychological safety and confidence in knowledge on employee knowledge sharing. *Manufacturing & Service Operations Management*, *11*(3), 429–447.

Stempe, J., & Badke-Schaub, P. (2002). Thinking in design teams-an analysis of team communication. *Design Studies*, *23*(5), 473–496.

Sutton, R. I., & Hargadon, A. (1996). Brainstorming groups in context: Effectiveness in a product design firm. *Administrative Science Quarterly, 41*(4), 685–718. http://www.jstor.org/stable/2393872. ISSN: 00018392.

Talton, J. O., et al. Exploratory modeling with collaborative design spaces. *ACM Transactions on Graphics-TOG, 28*(5), 167.

Tohidi, M., et al. (2006). Getting the right design and the design right. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1243–1252). ACM.

Wauck, H., et al. (2017). From in the class or in the wild? Peers provide better design feedback than external crowds. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*.

Xu, A., et al. (2015). A classroom study of using crowd feedback in the iterative design process. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1637–1648). ACM.

Xu, A., Huang, S.-W., & Bailey, B. (2014). Voyant: Generating structured feedback on visual designs using a crowd of non-experts. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 1433–1444). ACM.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*(5), 459–482.

Yu, L., & Nickerson, J. V. (2011). Cooks or cobblers?: Crowd creativity through combination. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1393–1402). ACM.

Yu, L., Kittur, A., & Kraut, R. E. (2014a). Distributed analogical idea generation: inventing with crowds. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1245–1254). ACM.

Yu, L., Kittur, A., & Kraut, R. E. (2014b). Searching for analogical ideas with crowds. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems. CHI '14* (pp. 1225–1234). Toronto, Ontario, Canada: ACM. https://doi.org/10.1145/2556288.2557378. ISBN:978-1-4503-2473-1.

Yu, L., Kittur, A., & Kraut, R. E. (2016a). Distributed analogical idea generation with multiple constraints. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. CSCW '16* (pp. 1236–1245). San Francisco, California, USA: ACM. https://doi.org/10.1145/2818048.2835201. ISBN:978-1-4503-3592-8.

Yu, L., Kittur, A., & Kraut, R. E. (2016b). Encouraging "Outside-The-Box" thinking in crowd innovation through identifying domains of expertise. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. CSCW '16* (pp. 1214–1222). San Francisco, California, USA: ACM. https://doi.org/10.1145/2818048.2820025. ISBN:978-1-4503-3592-8.

Zyskowski, K., et al. (2015). Accessible crowdwork?: Understanding the value in and challenge of microtask employment for people with disabilities. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1682–1693). ACM.

Towards Empirical Evidence on the Comprehensibility of Natural Language Versus Programming Language

Patrick Rein, Marcel Taeumel and Robert Hirschfeld

Abstract In software design teams, communication between programmers and non-programming domain experts is an ongoing challenge. In this communication, source code documents could be a valuable artifact as they describe domain logic in an unambiguous way. Some programming languages, such as the Smalltalk programming language, try to make source code accessible. Its concise syntax and message-passing semantics are so close to basic English, that it is likely to appeal to even non-programming domain experts. However, the inherent obscurity of technical programming details still poses a significant burden for text comprehension. We conducted a code-reading study in form of a questionnaire through Amazon Mechanical Turk and SurveyMonkey. The results indicate that even in simple problem domains, a simple English text is more comprehensive than a simple Smalltalk program. Consequently, source code in its current text form should not be used as a reliable communication medium between programmers and (non-programming) domain experts.

1 Introduction

Software can generate value in many different domains whose experts are not necessarily programmers. Thus, the creation and maintenance of such software for domain-specific projects leads to collaboration of domain experts and programmers. Both parties benefit from frequent communication and knowledge exchange. On the one hand, experts need to articulate the details they expect programs to do such as convenient data collection, processing, and visualization. On the other hand, programmers

P. Rein (✉) · M. Taeumel · R. Hirschfeld
Hasso Plattner Institute for Digital Engineering, Campus Griebnitzsee,
14482 Potsdam, Germany
e-mail: patrick.rein@hpi.de

M. Taeumel
e-mail: marcel.taeumel@hpi.de

R. Hirschfeld
e-mail: robert.hirschfeld@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_7

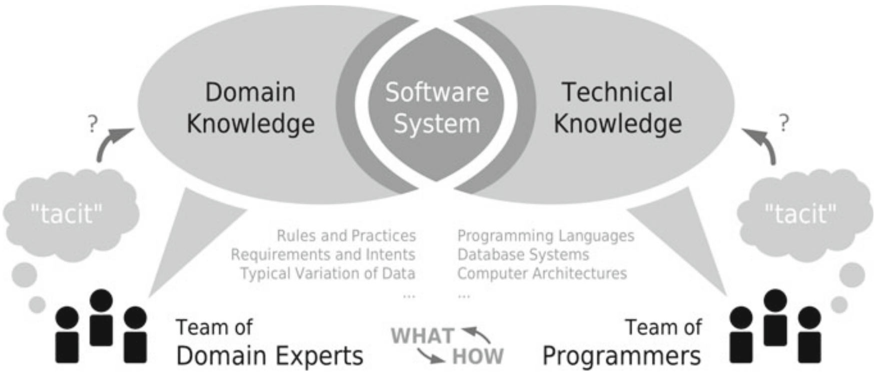


Fig. 1 An illustration of the way domain experts and programmers exchange knowledge when designing software systems

need to articulate technical possibilities and challenges to guide (and sometimes constrain) the experts' expectations. Eventually, such a symbiosis of two professions can yield the creation of something more valuable than the sum of its individual contributions (see Fig. 1).

In practice, programmers and domain experts try to establish a *shared language* to reduce the cognitive overhead in their discussions. Given an English-speaking team, such language might be regarded as "project-specific English" or "domain-specific English". The use of a shared vocabulary can reduce the number of comprehension errors and improve the team performance. This way, domain experts are *required* to express the relevant properties and rules in consistent terms. Programmers should apply these terms during code-authoring tasks. Given that a *software system* is being designed, we made two major observations: (1) all *words* in this language must have a *consistent meaning* and (2) *expressions* (or phrases) in this language have to be (made) *executable* for computers.

One approach to achieve such a shared communication medium is the use of (domain-specific) source code. Programmers are able to design domain-specific documents that still have meaning for the computer. Hidden technical layers can make the visible source code *look like* it was written in a (non-technical) domain language. Familiarity of (visual) form is often the basic argument for why domain experts might be able to think "in code". Eventually, there will always be complex, low-level details in the source code, but experts are not required to ever see them.

There are many domains where *natural-language text* plays an important role for capturing project-specific details. In such domains, the application of source code as a communication medium seems straightforward. While there might be technical details that distract non-programming experts, *domain-specific vocabulary* is discoverable in the form of high-level, executable code expressions (Evans 2004). This assumption leads to our research question:

> How can program source code be used as a frequent communication medium when exploring (or discussing) domain-specific terms and rules, which can also be expressed as natural-language text?

Generally, when using source code as a communication medium, two opposing factors might affect the comprehensibility. The first factor is that domain-specific source code makes aspects of the domain knowledge explicit. For example, types of relevant objects of the domain, relationships between objects, or rules of the domain all have to be made explicit in order for source code to become executable. This explicit description of knowledge *reduces ambiguity* and creates a *detailed description of domain knowledge*. In contrast, even domain-specific source code documents inevitably include the formal syntax and the complex semantics of the programming language. For a reader with little to no programming experience, these might impose an *additional complexity burden* when trying to understand even simple code segments.

We suspect this additional complexity burden to outweigh the benefits of a less ambiguous description of domain knowledge, at least for non-programming readers. We see two implications if our claim holds. First, given the current means for designing domain-specific programming languages, there is still a major need for *external documentation* and thus other (maybe non-executable) forms of representing domain knowledge. Second, there is a significant value for language researchers to explore interactive means for programming beyond textual languages.

We present the results of a user study based on a controlled experiment on the assumption that even domain-specific, object-oriented source code poses comprehension challenges for non-programmers.

We chose the object-oriented paradigm because real-world domains can easily be modeled as object collaboration (Evans 2004). We chose Smalltalk (Goldberg and Robson 1983) as a representative for an object-oriented programming language because its grammar is *close to basic English*. There, simple phrases consist of subject, predicate, and object. In Smalltalk, there are also objects that collaborate by sending messages. In comparison to English, the Smalltalk receiver of a message send is a *grammatical subject*, the Smalltalk message itself forms the *grammatical predicate*, and any message payload can be seen as *grammatical objects*. For example, the scenario "Every Friday, the postman delivers some mail to my address" translates to: "Date today dayOfWeek = #Friday ifTrue: [aPostman delivers: someMail to: myAddress]". In this way, Smalltalk can easily be applied to hide technical details such as request handling and resource management.

The hypothesis to test in this work reads as follows:

> Given a problem domain with simple rules, people with little to no programming experience understand less details from a Smalltalk program than from an English text document.

Here, we define "simple" as (1) in the order of 10 domain rules, (2) in the order of 5 kinds of objects (or concepts), (3) less than 300 lines of Smalltalk code, and (4) less than 500 words English text. We base these figures on personal experiences from past projects. We conducted the experiment with 31 participants. As a result, we found that while participants can answer questions about scenarios expressed in source code documents, they answer more questions correctly when working with a corresponding natural language document.[1]

In Sect. 2, we describe our experimental setup in detail by describing the layout, tasks, and the participants. Our user study is a questionnaire, which we carried out through Amazon Mechanical Turk (MTurk) and SurveyMonkey and whose implementation we describe in Sect. 3. In Sect. 4, we present and discuss our quantitative and qualitative results. In Sect. 5 we give an overview of related work in terms of text and code comprehension. Finally, Sect. 6 presents concluding thoughts and we outline future work.

2 Experimental Design

As our experimental design is not based on a previous design, we first describe our experimental layout in detail including the operationalizing of the relevant variables. We further describe the design of the scenarios used in the comprehension tasks, as well as the design of the used questionnaires. As we conducted the experiment through MTurk, we report on the measures we took to ensure that we got participants with the target background.

2.1 Experimental Layout and Operationalizing of Variables

We conducted a quantitative user study through a controlled experiment with a fixed setup (Robson 2002). Our hypothesis addresses the *difference in comprehension* for two forms of text documents: (1) natural language and (2) programming language. Thus, we had to *operationalize* the variables (a) text comprehension and (b) programming experience. Note that we target English-speaking readers with *little to no* programming background. This renders such an experimental design very challenging in terms of transparency and reproducibility. Consequently, we employed a *within-subject* layout to at least cope with variations among different participants and focus on mitigating carry-over effects between rounds for individual participants. We designed *two similar tasks*, namely a conference-registration process and a store-checkout process, whose order we counter-balanced, including the text-or-code treatments. On the downside, we require sets of four participants to equally serve all groups (as depicted in Fig. 2):

- Group A: Round 1 is Text (Registration), Round 2 is Code (Check-out)
- Group B: Round 1 is Text (Check-out), Round 2 is Code (Registration)
- Group C: Round 1 is Code (Registration), Round 2 is Text (Check-out)
- Group D: Round 1 is Code (Check-out), Round 2 is Text (Registration)

[1]You can access the archived materials used for conducting the experiment and the unprocessed and pre-processed data from https://doi.org/10.5281/zenodo.2540989.

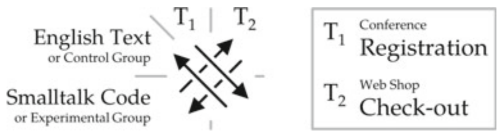


Fig. 2 The overall experiment layout showing all four groups

Having such a setup, we depend on proper *training* to familiarize all participants with our notions of *text document*, *English style, and programming style.* In addition to learnability, we do not address other carry-over effects explicitly.

One could argue that a participant's *motivation* might affect the upper time limit of a round. At the same time, we expect participants to unnecessarily rush through rounds, which might make a lower time limit more relevant. As we had no prior experience with similar experiments, we did not impose any time limits.

To collect measurable results, we operationalized *text comprehension* through *answers in a questionnaire* about the document's content. Thus, we take the number of correct answers as a measure of the level of comprehension. For content presentation, we chose English as the natural language to attract many potential participants. We chose Smalltalk as the programming language, because of its small syntactic overhead (Goldberg and Robson 1983). Source code can be expressed in a concise fashion, sometimes almost matching their English-reading explanation.

To discard participants early in the process, we operationalize *programming experience* as the *number of lines of code written in the past* reported through *formal self-evaluation*. We expect this number to be very low or zero at best. Also, we query their experience with major programming paradigms such as the object-oriented paradigm (Wegner 1987) or the functional paradigm (Abelson 1996) on a five-point Likert Scale (Robson 2002).

Finally, we want to stress our goal of *simplicity* in our experimental tasks (or scenarios). On the one hand, we favor simple English text. It should be neither literarily verbose nor artificially instructive. On the other hand, we favor simple program code in a clean style, closely resembling the vocabulary of the task's domain (i.e., conference registration or Web shop check-out). Consequently, our experimental design does not investigate isolated features of the documents but only the difference between a particular style of *natural-language text* and a particular style of *programming-language code*. To inform the creation of more detailed hypotheses, we collected *qualitative data* on the aspects of the documents the participants' thoughts and struggles in the debriefing phase.

2.2 Scenarios: Creation Process and Document Properties

Participants complete two rounds back-to-back. First, they answer questions about one scenario described in one language. Second, they answer questions about another (similar) scenario described in another (dissimilar) language. For each scenario, there are *ten questions* about the content and *two control questions* to verify participant authenticity, that is, to filter mechanical (or unserious) behavior. We do not impose time limits on the tasks.

Both scenarios are *representative* in the sense that they model existing, software-supported processes (or systems). In one scenario, we describe a conference-registration process, adapted from a real-world example one of the authors worked on. In the other scenario, we describe a shop-checkout and delivery-planning process, which has common characteristics of today's online-shopping processes.

Each scenario's domain guides the vocabulary, concepts, and rule of the descriptions (and questionnaires). The conference-registration process describes the online registration of a conference. It describes the interactions with prospective conference *visitors*, rules for deciding which extra *services* they can book, rules for deciding whether they can be registered or are put on a *waiting list*, and special cases for visitors with *invitation codes*. The shop checkout process describes the interactions with a customer of an online store for *books*, rules for deciding whether a customer is eligible for certain *shipping modes*, and a procedure for determining a *time of delivery*.

Both scenarios are similar yet different. They are similar in terms of their extent and structure, which addresses number of phrases (or code lines) and the choice of (English or Smalltalk) syntax. Yet, they are different considering their *source of complexity.* While, the registration process is simple with a complex set of rules for the visitor, the check-out process is more complex with a simple set of rules for the delivery slot. We discussed the comparability of tasks in Sect. 4.4 as one source of threats to validity in our experimental setup.

To ensure similar structure, we first wrote the natural-language text and only then the corresponding code. After code authoring, we iterated over the English description, again, to align any phrases that turned out to be difficult to express in Smalltalk. This iterative procedure revealed unnecessary verbosity in the English text and artificial abstractions in the Smalltalk code. We formulated the questionnaire afterwards.

The style of writing can influence the comprehensibility of the documents. We found several rules to guide the application of the English language and the selection of Smalltalk features. While our main goal was to increase similarity between document representations, it helped us maintain a simple, yet representative, baseline. The rules for our English texts are:

- **Overview before detail**: Both scenarios describe processes. Thus, the reader should grasp the entire scope before diving into process-step details.
- **Paragraphs represent single steps and rules**: Readers might have to look up details in the description repeatedly as questions are answered. That look-up can be simplified if process steps and rules are at the granularity of text paragraphs.
- **Alternatives begin with "if"**: There are conditions, which divide the processes into branches to be understood and recalled frequently. Thus, the reader should recognize such branches directly at the beginning of such sentences, not in the middle.

- **No synonyms**: Once decided on characteristic vocabulary for each scenario, the reader should not be bothered with extra complexity through synonyms. We kept a consistent vocabulary throughout process descriptions.
- **No text emphasis**: Being one characteristic difference between natural language and programming language, we reserve elaborate emphasis (such as colors, bold face, and italic face) for source code representation, not English text.

The rules for our Smalltalk code are:

- **Objects first**: Since Smalltalk favors the object-oriented programming paradigm, we model domain concepts with objects (and object classes) as much as possible.
- **Descriptive identifiers**: The vocabulary used to form Smalltalk expressions should reflect the domain as much as possible. This includes identifiers for variables and messages (or method names).
- **No explicit loops**: We want to avoid the cognitive effort of understanding loop constructs in code (Robins et al. 2003). Smalltalk offers a concise alternative, similar to functional languages, with its collection protocol in the form of messages such as #select:, #count:, #firstThat:.
- **No recursion**: We want to keep the concept of message look-up as simple as possible. Thus, we want to avoid the idea of methods calling themselves repeatedly.
- **No meta-programming**: Smalltalk systems offer elaborate means to monitor and modify themselves. This powerful mechanism adds cognitive overhead that is not required for expressing simple domain rules. This includes the use closures in anonymous functions.
- **Descriptive data access**: For each data field, use an eponymous message (i.e. "*field*") to read but "*changeFieldTo:*" to write that field value. While this is against best practices in Smalltalk, pilot runs indicate that "*field:*" is too difficult to understand.
- **Minimize comments**: Comments reflect any additional information not represented in the code directly. None of the comments contain relevant facts for the questions.

Considering code formatting, we want to primarily guide document navigation in the experiment. We are aware that this is different from how programmers consume code in integrated programming environments. In particular, we use color to highlight structure and guide message look up. We highlight identifiers that store a value with a class described in the task in *green* and methods that can be looked up in *blue*. Additionally, code for each class is set in a section starting with a *distinct heading*. Each method is set as a distinct section (see Fig. 3). Overall, the code is formatted according to a standard set of idioms (Beck 1996). As an example, English text phrases correspond to Smalltalk code statements.

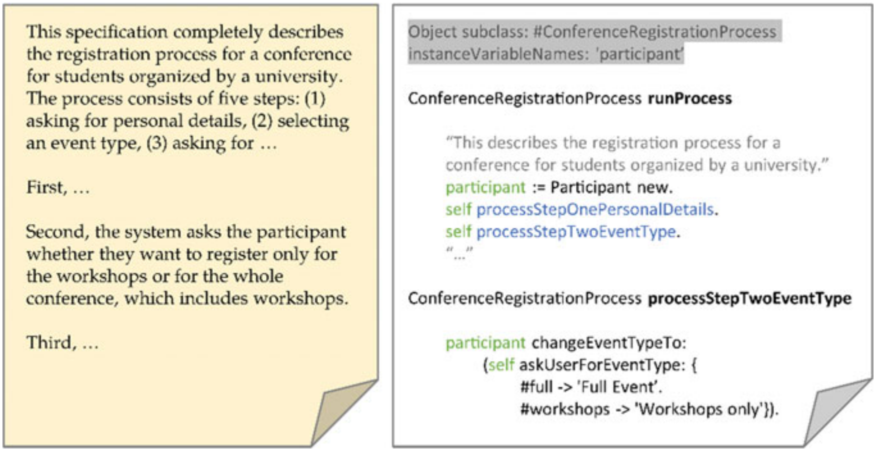


Fig. 3 A comparison of a textual section and the corresponding source code section, including the actual type setting of the source code documents used during the experiment

2.3 Questionnaires: Structure and Measurement

We designed questionnaires to gather *quantitative data* about the level of comprehension and *qualitative data* about general feedback on our experimental setup. There are *single-choice* questions on the scenario details and *free-text* questions on general concerns.

We distinguish two kinds of comprehension questions: (1) facts and (2) applications. First, questions about facts cover general properties and rules of the respective process. Such questions usually begin with "In which of the following circumstances can the following happen?" Participants can read these facts directly from the documents. Second, questions about applications require participants to assume particular properties for processes in action and derive implications. Such questions are usually like, "Given X, which of the following can happen?" or begin with "For this question, imagine the following scenario: ...". Participants must therefore simulate the respective effect in their mind.

All comprehension questions are, *virtually,* single-choice questions. That is, even if there is more than one correct answer, participants are only required to pick one. We think that this is a fair trade-off for our hypothesis because multiple interpretations might be valid for vague (or abstract) descriptions. Note that for each question, we added two answers that express *generic* comprehension issues:

- Nobody can read this fact out of the given materials.
- I cannot answer the question.

Participants can therefore indicate that the provided materials do not provide enough information on some concern. We think that it is crucial in communication to be able to consciously express "I don't know" and guess "Nobody can know".

This indicates that the current version of a document is not clear enough to serve as baseline for knowledge exchange between people. Such documents are usually created and maintained in an iterative manner. Mistakes can happen, waiting to be corrected.

For the (qualitative) feedback questions, we wanted to collect as much additional insight on our experiment and hypothesis as possible. Since we use MTurk, we cannot interview and debrief participants in person. Thus, an emerging structure of interview questions is not possible. Still, we encouraged participants to quote relevant passages from the provided documents to help us understand and interpret their concerns. We suggested the following topics for feedback: length of tasks, difficulty of questions, and obstacles to understanding.

In the end, a high level of understanding means a high number of correctly answered questions. We are aware that our selection of questions might not capture the entire knowledge base the documents encode. Yet, we did not consciously introduce irrelevant facts to distract or confuse participants.

2.4 Participants: Selection and Training

Our hypothesis addresses a wide range of people working in any field, but who do not have elaborate programming skills. Thus, we employed volunteer sampling by recruiting participants through MTurk to get a sample set of our target group. Since we cannot personally interview participants upfront, we installed several automated guards to ensure authenticity and high-quality results:

- **Basic engagement**: We required participants to have the "Amazon Mechanical Turk Master Worker" qualification, which indicates consistent high-quality work in a variety of tasks with a variety of contractors (or domains?).
- **Basic skills**: We required participants to have a U.S. bachelor's degree, because we aim for scenarios that involve software developers and domain experts. We want to investigate their ways of knowledge exchange.
- **"Bot" detection**: We inserted control questions to check for the participant's continuing attention. This might also reveal automatic (or unserious) behavior of participants who just want to collect the reward through MTurk very quickly.

We verify the authenticity of participants with *four control questions*, which we insert randomly throughout the questionnaires. Each question checks for common knowledge or basic involvement with the provided material. They are as follows:

- Which of the following kinds of processes does the document describe?
- Which of the following numbers is less than 100?
- What kind of item does the store sell?
- Which one of the following things is generally considered the tallest?

Even if participants are genuine, engaged, and educated, they cannot know what "little programming skills" means. They might accept this task erroneously, but with good intentions. Thus, we operationalized *programming experience* as a formal self-evaluation using questions to query written lines of code and familiarity with common programming paradigms. We rejected participants who reported *more than 1000 lines of code*. If such a number cannot be estimated, we resort to paradigm familiarity on a five-point Likert Scale: strongly agree, agree, neither agree or disagree, disagree, or strongly disagree. We provided the following statements to assess:

- I am familiar with the logic programming paradigm.
- I am familiar with the functional programming paradigm.
- I am familiar with the object-oriented programming paradigm.
- I am familiar with the Smalltalk programming language.

We think that familiarity with the Smalltalk language can also bias our results. Yet, we do not expect to find many Smalltalk programmers through Amazon Mechanical Turk. We rejected participants who expressed that they "strongly agreed" or "agreed" to any statement.

As we assume that working with formal models other than programming languages (such as construction processes and rule sets, chemical equations, statistical analysis, laws to some extent) might have a substantial influence on the performance on our tasks, we recorded the participants' experience with formal methods, and their professional and educational background. We did not reject participants based on this information.

With these filters in place, we recruited 35 participants with the expectation of achieving 30 genuine submissions. We rejected one participant due to level of programming experience, and ultimately recruited 36 participants, gaining 31 genuine submissions.

Since our experiment design entails potentially unfamiliar tasks, we provided *training tasks* to level relevant previous experience with reading comprehension tasks and thus mitigate carry-over effects from learning between rounds. This training consisted of an example scenario and questionnaire for the participants to practice. There was a shorter, but similar, textual description and *two* questions that also had the same structure as the questions used in the actual experiment.

As we explicitly aimed for readers with little to no programming experience, we did not provide training on the *semantics* of source code. However, we provided basic guidance on the *structure* of source code documents. Participants have to look up information, which might be straightforward in natural-language text but novel and uncommon in source code. So, we explained the role of *identifiers and method selectors* and our use of color for simplification. By example, we described this generic approach to find and connect information in code documents.

3 Experiment Procedure

In this section, we describe the experiment procedure. We recruited participants through MTurk. The screening of participants and the actual experiment run was implemented using the SurveyMonkey service. After these automated processes, we manually filtered and analyzed the participants' results.

3.1 Participant Management in Amazon Mechanical Turk

We posted the experiment as a task on MTurk, following guidelines for *academic requesters* written by the "Dynamo MTurk Community".[2] These guidelines cover fairness and respect for participants including authenticity, privacy, payment, or time (or work) estimate. We advertised our experiment with the keywords "study" and "text comprehension" as follows:

> Answer questions about complex processes described in text and code (limited to participants with little to no programming experience) (~90 to 120 min).

The estimated duration of the experiment is provided to give participants an impression of the expected time investment. We have no means to enforce this estimate. The participants were reimbursed US$15 on completion of the task or a fraction of this amount if they exited the task earlier. See Sect. 2.4 for more details on our selection criteria. Note that we excluded all participants of (prior) pilot runs of the experiment through a project-specific MTurk qualification.

When previewing the MTurk task, participants were able to see the terms of the experiment. Unfortunately, we do not know how many participants saw the *initial* description including the reward or the preview of the terms. When they accepted the task, they were provided a web link to the SurveyMonkey questionnaire. At that point, our tracking of participants starts.

At the end of the SurveyMonkey questionnaire, participants got a completion code, which they entered on the MTurk task page to submit the task and eventually earn the reward.

3.2 Questionnaire via SurveyMonkey

We conducted the main part of the experiment through a SurveyMonkey questionnaire. We collected both *complete and incomplete* sets of answers because SurveyMonkey records all started sessions in the same database. Such incomplete records occur when participants exit the experiment early. They could decide to quit at any time or be disqualified by the process automatically during the initial screening (or skill-check) questions.

We designed the questionnaire for separate pages (for an overview see Fig. 4). The

[2] See https://web.archive.org/web/20190115201202/http://wiki.wearedynamo.org/index.php/Guidelines_for_Academic_Requesters and https://web.archive.org/web/20190115201354/http://wiki.wearedynamo.org/index.php/Basics_of_how_to_be_a_good_requester).

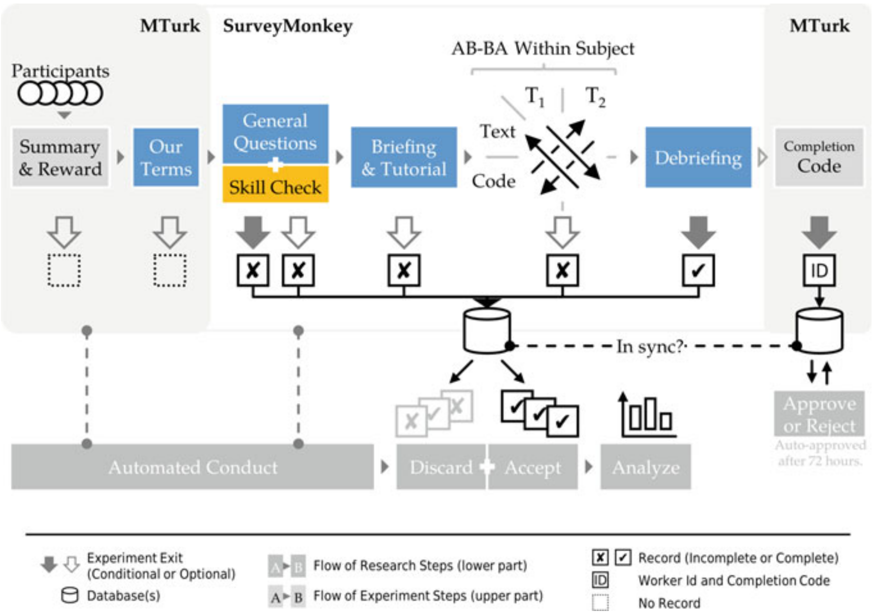


Fig. 4 An overview of the procedure of the experiment. The upper section shows the actual run of the study with the participants. The bottom section shows the overall process including pre-processing of data and the data analysis

first page asked *general questions* about gender, age, and professional background. This page also included the questions related to participants' *programming experience*. Depending on the answers, participants were automatically disqualified (see Sect. 2.4), for example if someone was an expert in a certain programming paradigm. After this screening, participants received a *general briefing* on the structure of the questions and documents. That is, we suggested a primary reading method for both code and text. After this briefing, there was a tutorial page for training (see Sect. 2.4). It contained example questions through which participants could familiarize themselves with the types of questions used throughout the experiment. This training concludes the preparation.

Participants then started with the actual experiment. At the beginning of each questionnaire, we added weblinks to the description of the current scenario as well as the general description on how to read (code) documents. We *disguised* these weblinks to prevent participants from peeking at the scenario's other representation, which was code or text respectively. For example, one document was located at "…/code-fhbdz.html" and the other at "…/text-129d.html". Further, we used the SurveyMonkey mechanisms to implement the randomized assignment to participants of (a) document types, (b) scenarios, and (c) ordering. We further used SurveyMonkey to randomize the ordering of questions for each scenario as well as the ordering of single-choice options for each question.

The questionnaire concluded with a debriefing page, which included a question on general remarks as well as the completion code for MTurk to collect the reward.

3.3 Pre-processing the Results

As described above, we put several guards in place to ensure participant authenticity and thus genuine results. However, after the experimental conduct, the databases can include also records with obviously strange properties. Such records would unnecessarily bias statistical analyses and thus our interpretations. We employed the following steps to clean those records:

- **Synchronization with MTurk**: For all records on SurveyMonkey, we checked whether the provided MTurk worker identification number was listed in the list of workers who accepted the work on MTurk. We discarded all submissions whose identification numbers were not listed.
- **Repeated submissions**: We compared the IP addresses from which the sessions on SurveyMonkey were conducted. If the same IP address showed up in two records, we examined further for authenticity. That is, two people could share the same Internet connection—which would be acceptable. Otherwise, we discarded those submissions.
- **Incomplete submissions**: We discarded submissions that did not answer *all* mandatory questions correctly or incorrectly. Note that there were also feedback questions to gather qualitative data in the form of free-text fields, which were not mandatory.
- **Control questions**: We discarded submissions that did not answer all control questions *correctly*. We assume that such participants did not put the required effort into the experiment.

Note that we did not enforce time constraints. However, we noticed several records indicated low effort in terms of minutes spent. While we have no reason to remove them from statistical analyses now, we should think about ways to increase time spent and effort invested in follow-up studies.

4 Results and Discussion

After pre-processing the data, we analyzed the results with regard to the initial hypothesis. To ensure that our assumptions about the background of our participants were met, we determined several characteristics of our group of participants. Beyond the analysis of the quantitative results, we also summarized the results from the qualitative feedback we collected. Finally, we discussed threats to the internal and external validity of our results.

4.1 Participants' Background and Behavior

As we used volunteer sampling, our sample is not a true random sample. Thus, before describing the results of the comprehension tasks, we will first describe the characteristics of the participants. As a result of the described screening of participants and pre-processing of submitted questionnaires, we ended up with results from 31 participants. The participants are not equally distributed among groups (see Table 1). This is since many participants signed up before a single participant had finished the questionnaire. Thus, we could not balance groups after participants dropped out or failed control questions.

Regarding the programming experience of the participants, most participants self-reported that they had never written any source code (see Table 2). Three of the five participants who reported that they wrote some code (5–1000 lines of code) stated that they wrote small scripts to configure user interface scripting systems. The other two participants did not state what kind of code they created.

Except for one, all participants either selected "disagree" or "strongly disagree" for any of the statements about their familiarity with any of the programming paradigms, the familiarity with the Smalltalk programming language, or their usage of formal methods. One participant selected "neither agree nor disagree" for two of these questions.

Finally, we determined the time spent on the questionnaire by taking the difference between the time the SurveyMonkey session was started and the submission of the last page of the questionnaire (see Fig. 5). Most participants spent 24–39 min on the questionnaire. Generally, most participants stayed under 99 min in total (excluding one participant whose recorded time was 22 h and 58 min which was due to a technical issue on the side of the participant). As we intended participants to spent between 60 and 120 min on the task, the time participants actually spent on the task was lower than we intended.

Table 1 Number of participants per experimental group

Group	A	B	C	D
Number of participants	8	9	5	9

Table 2 The number of participants who reported a certain number of lines of code they wrote in the past

Reported LOC	0	5	10	100	1000
Number of participants	26	1	2	1	1

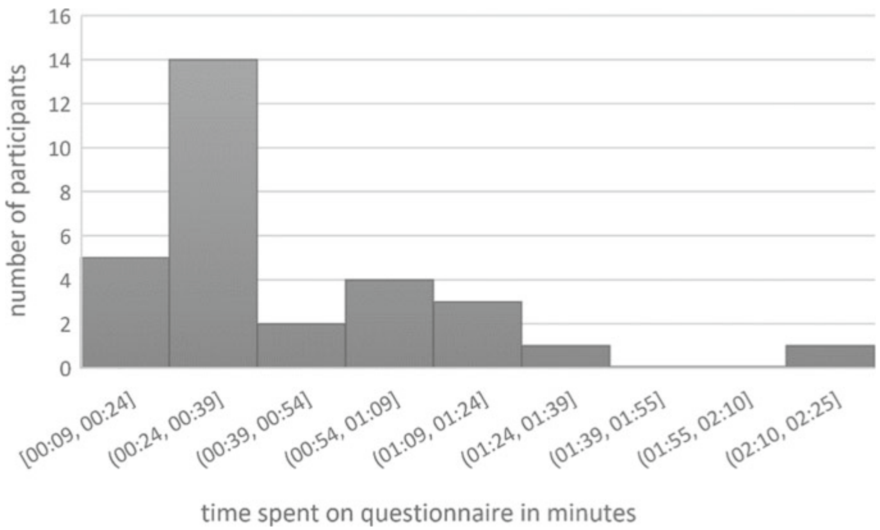


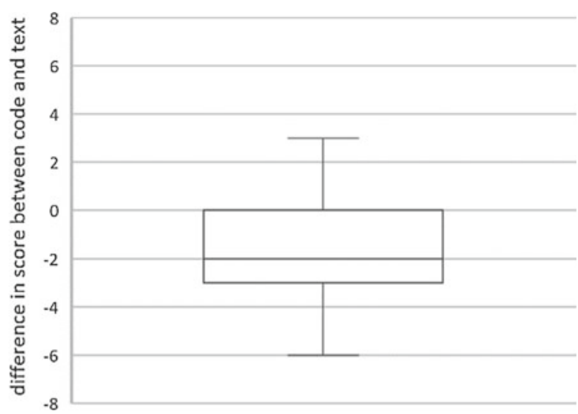
Fig. 5 Histogram of the time spent on the questionnaire excluding an outlier of 22 h due to technical difficulties of the participant. Each compartment spans an interval of 15 min

4.2 Analysis of Quantitative Results

The primary goal of the experiment was to determine whether the format of the document influenced comprehensibility (see Fig. 7 for an overview of the results). As we employed a within-subject design, we used a paired t-test for comparing the results. Therefore, we determined the mean of the differences between the scores for the Smalltalk and the English language document of each participant (see Fig. 6).

The differences were normally distributed, as assessed by the Shapiro-Wilk's test ($p = 0.197$). The mean difference is -1.42 (standard deviation 1.747) and the

Fig. 6 A box plot of the distribution of the score differences for each subject between the code and the text document



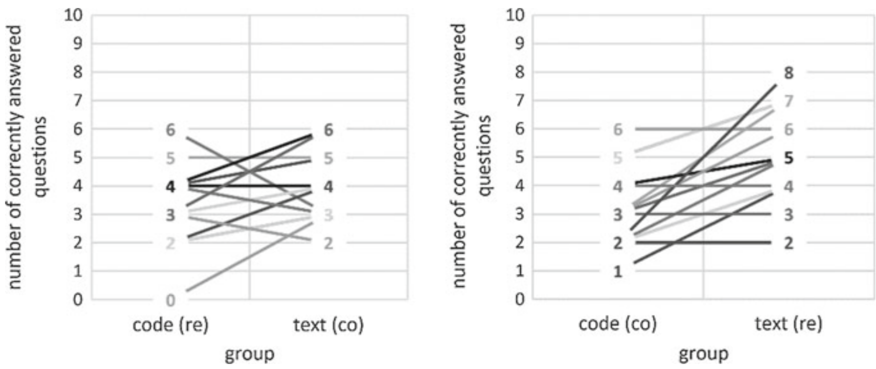


Fig. 7 Individual results for each participant between the questions answered on code and the questions answered on a text document. The two diagrams show the results for the respective combinations of document formats and scenarios (re = registration process, co = shop checkout)

difference between the scores is significantly different from zero ($t_{31} = 4.524$, $p < 0.001$). This difference means that on average participants scored fewer points on questionnaires about a Smalltalk document than they did on questionnaires about a text document.

Further, we designed the scenarios and the corresponding questionnaires to be of equal difficulty, to prevent any differences in difficulty from potentially influencing the results. To check this assumption, we also analyzed the difference between the scores for the conference registration and the shop checkout scenario. Again, the differences were normally distributed, as assessed by Shapiro-Wilk's test ($p = 0.356$). The mean difference between the two scenarios is not significantly different from zero ($t_{31} = 1.656$; $p = 0.108$).

At the same time, the plots of the two groups show that the groups do in fact differ. For example, the group that worked on the registration process expressed in code had a lower minimum score for the code task and a lower maximum score for the text ask. Further, this group had two outliers who performed better on the code document than on the text document.

Despite the statistically significant results, the results should be regarded as preliminary. A detailed analysis of variance would be required to clarify the influence of the scenarios onto the results. Additionally, the current analysis does not cover the influence from the time spent on the questionnaire, the ordering of scenarios, or any interactions between these variables.

4.3 Summary of Qualitative Feedback

Following the completion of each questionnaire about a scenario, we asked participants to describe any difficulties they encountered when working with the documents.

We did not apply a rigorous analysis to the data but summarized some of the themes in the responses below. As the participants wrote these statements after they finished all questions on one scenario, the statements do not allow for any conclusions on the actual difficulties the participants encountered. Nevertheless, they can serve as a starting point for future programming language features or experiment setups.

One topic which was primarily mentioned with regard to working with code documents was that participants felt that some *statements in the code lacked relevant context*. One participant stated this quite explicitly:

> […] the apparent answer to the previously stated issue appears later in this code block, but not in direct proximity to it, which may result in transient confusion.

The confusion was related to the meaning of the value of a variable in the conference registration process. Although the value of the variable was "#registered" the meaning of the value with regard to the registration process was unclear until the value was used later on to notify users that they were successfully registered. Another participant stated that they were unsure about the definition of a certain group of conference visitors:

> […] I also didn't understand if people who wanted to participate in the conference were also counted as going to the workshop and if that limit therefore applied to them (the 100 people limit). So, this line was challenging: 'self numberOfWorkshopParticipants >= self limitOfWorkshopParticipants' […]

In order to understand the quoted code statement, one would have to look up the definition of the two identifiers "numberOfWorkshopParticipants" and "limitOfWorkshopParticipants". Both identifiers were defined in the code but in a different section of the code document. Again, the spatial distance between the statement and the corresponding detailed definitions might have caused the confusion.

A second topic with regard to the code documents was *difficulties with technical vocabulary or syntax*. While we intended to keep the code free from any unnecessary technical details, some details remained in the code. Some of these were mentioned as confusing or completely obscure:

> […] I did not understand what this meant: notNil so I could not interpret that code line at all. […]

The "notNil" statement was also mentioned by other participants. However, only the quoted participant stated that this was a hindrance to understanding the surrounding line. Others simply noted that they were not sure what it meant. Another technical aspect of the code that was mentioned was the hash character, which is the syntax for denoting a symbol in Smalltalk (a piece of text with a unique object identity). An example in the source code is the symbol "#registered". This syntax was mentioned by one participant who quoted a section of code and stated that they were unsure about the meaning of the hash character:

> […] 'ifFalse: [| resultStatus |
>
> resultStatus : = #registered'but I don't understand what the # means. […]

Notably, these two were the only technical aspects mentioned although the code did include a variety of technical vocabulary and syntax, such as the method name "->", square brackets, or curly braces.

Besides these topics primarily related to the code documents, one interesting theme with regard to the text documents was that participants would have preferred a different layout or representation of the text:

> It was hard to apply the information in paragraph format. I had to break it down line by line. […]
>
> The challenge, for me, is that the information wasn't broken up into paragraphs or bullet points.

One participant even went so far as to suggested that a graphical representation of the process logic might help with understanding the underlying structure.

> […] I think in terms of my understanding and answering, some sort of flow chart or similar graphic might be the most useful. This would allow you to sort registrants methodically step by step based on attributes like local/visiting, code/no code, workshop-only/full […]

4.4 Threats to Validity

While we consider the results of the analysis preliminary, we also want to point out the particular threats to the internal and external validity we identified with the current setup.

One *internal threat* is the fact that a large number of participants spend less than the intended time on the questionnaires. As the scenarios are complex and we can assume that the questions require concentration, spending less than 30 min on the tasks suggests that participants did not fully engage in the tasks. This matches the feedback on the overall study of one participant, who said: "I expected this to be a lot longer study. I would possibly include a progress bar […]". Overall this might mean that we did not measure how well the documents might be comprehended but how well they might be skimmed or how quickly readers could find information relevant to a question.

Another *internal threat* results from conducting the experiment on MTurk. As we have no way to control the activities of participants while participating in the experiment, participants might have used external data sources for understanding the Smalltalk code, for example by searching for the meaning of method selectors. While we cannot control for it in general, future designs should incorporate a question at the end of the experiment as to whether participants used any information not contained in the experiment material.

As mentioned in the analysis of the questionnaire scores, the comparability of the difficulty of tasks might still be a challenge and thus be an *internal threat.* Additional analysis is required to determine the differences resulting from the scenarios. If the analysis shows that their difficulty differs, future designs will have to balance them better.

Despite two pilot runs, overlooked technical details in the scenario descriptions, such as the message "notNil" (see Sect. 4.3), might have confused participants unintentionally and thus be another *internal threat*.

An *external threat* is the limitation of features of the Smalltalk programming language we used in the code documents. While Smalltalk code generally contains only few explicit loops, these might be necessary for some domain logic. At the same time, one could argue that if the domain logic does require an explicit loop, the complexity introduced by the loop was inherent to the domain. Beyond these situations, performance optimizations might require programmers to use more advanced language features in a code document.

Further, the minimal training on the semantics of source code might be unrealistic in practice and thus be an *external threat*. In situations in which domain experts regularly interact with programmers, one might argue that domain experts would quickly become proficient in the programming language.

5 Related Work

As the readability of source code affects not only source code as a communication medium but also the accessibility of source code for programming novices and the productivity of professional programmers, related work covers a variety of topics.

For example, a systematic series of controlled experiments showed that the different programming language syntax of existing programming languages has an impact on the accuracy of readers with no or little programming experience (Stefik and Siebert 2013). The studies used a limited set of features, based on the set of features programming novices would first encounter when learning to program. The results of these studies were further used to inform the design of parts of the Quorum programming language.

Another study set out to investigate the impact of a domain-specific language on the performance of programmers (Mogensen Ingibergson et al. 2018). The experiment design particularly targeted programming with a domain-specific language. The participants were all programmers who had some experience with the C++ programming language. While the results were inconclusive, the insights for designing code readability experiment could be used in future iterations of our setup.

Besides studies on the effect of textual languages, some studies examined the differences between textual and graphical representations of code. For example, one study investigated the impact of a textual and a graphical notation for regular expressions for readers with programming experience (Hollmann and Hanenberg 2017). Regular expressions are a domain-specific language regularly used by programmers for pattern matching, for example in text segments. The study found that readers could answer questions on the expressions faster when working with the graphical notation.

Finally, a more general, related argument comes from literature on end-user programming (Nardi 1993). While programming languages are formal languages, studies on end-user programming suggest that the formal nature of code does not have to be an obstacle for understanding code documents. People regularly use formal languages in their everyday life without recognizing them as such. Examples for such languages are calculating sports statistics or sewing or knitting patterns. While the underlying languages are formal, they are not perceived by their users as such because they are primarily specific to a task that is familiar to its users.

Finally, a recent study suggests that the differences between code documents and text documents might become less distinct for experienced programmers (Floyd et al. 2017). The study investigated whether the brain activity while reading source code is more similar to reading mathematical texts or to reading prose. While the results are preliminary, they indicate that with higher expertise in programming, the brain activity seems to become more similar to reading prose.

6 Summary and Conclusions

In the described study we set out to gather initial empirical evidence on the assumption that even domain-specific, object-oriented source code is insufficient to express domain logic in a way accessible to readers with little to no programming experience at all. Therefore, we devised a design for a controlled experiment through Amazon Mechanical Turk comparing the comprehensibility of documents about domain processes expressed in either Smalltalk source code or the English language. A first run of the experiment design resulted in data from 31 participants. Despite having no or very little programming experience, most participants were still able to answer several of questions based on the code documents. Nevertheless, the data provides statistically significant results showing that code is less comprehensible. However, this result should be regarded as preliminary as further post hoc data analysis and more experiments are still required to clarify the influence of a potential difference in the difficulty of the scenarios.

The current experiment already hints that general-purpose programming languages might not yet be accessible enough. Beyond this, the described experiment setup could now be used for further experiments investigating how particular features of the code document or the natural language documents make the described domain knowledge accessible or not. These insights could then be used to design better languages and tools to make source code a useful artifact in the everyday communication between domain experts and programmers.

References

Abelson, H., Sussman, G. J., & Sussman, J. (1996). *Structure and interpretation of computer programs*. Justin Kelly.

Beck, K. (1996). *Smalltalk best practice patterns*. Upper Saddle River: Prentice Hall.

Evans, E. (2004). *Domain-driven design: Tackling complexity in the heart of software*. Addison-Wesley Professional.

Floyd, B., Santander, T., & Weimer, W. (2017). Decoding the representation of code in the brain: An fMRI study of code review and expertise. In *Proceedings of the 39th International Conference on Software Engineering*. IEEE Press.

Goldberg, A., & Robson, D. (1983). *Smalltalk-80: The language and its implementation*. Addison-Wesley Longman Publishing Co., Inc.

Hollmann, N., & Hanenberg, S. (2017). An empirical study on the readability of regular expressions: Textual versus graphical. In *2017 IEEE Working Conference on Proceedings of the IEEE Working Conference on Software Visualization (VISSOFT)*. IEEE.

Mogensen Ingibergson, J. T., Hanenberg, S., Sunshine, J., & Schultz, U. P. (2018). Experience report: Studying the readability of a domain specific language. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. ACM.

Nardi, B. A. (1993). *A small matter of programming: Perspectives on end user computing*. Cambridge: MIT Press.

Robins, A. V., Rountree, J., & Rountree, N. (2003). Learning and teaching programming: A review and discussion. *Computer Science Education, 13*(2), 137–172.

Robson, C. (2002). *Real world research* (2nd ed.). Malden: Blackwell Publishing.

Stefik, A., & Siebert, S. (2013). An empirical investigation into programming language syntax. *Transactions on Computer Education, 13*(4), 19. (IEEE).

Wegner, P. (1987). Dimensions of object-based language design. In *Proceedings of the International Conference on Object-oriented Programming, Systems, Languages & Applications* (pp. 168–182). ACM.

Team Creativity Between Local Disruption and Global Integration

Axel Menning, Benedikt Ewald, Claudia Nicolai and Ulrich Weinberg

Abstract What differentiates an average conversation from a creative conversation? In this book chapter, we answer this question by looking at coherence styles of design conversations. With the help of the Coherence Style Framework (CSF), we are able to illustrate what divergent and convergent thinking on the conversational level looks like. Highly creative teamwork is represented as an alternation between local disruption (local low coherence) and global integration (global high coherence). This has implications for the current practices of idea generation of design thinking and innovation teams.

1 Introduction

Teams are at the core of innovation and Design Thinking (Gilson et al. 2015; Valkenburg 2000). This is due to their ability to consider and synthesize multiple perspectives very effectively and efficiently, which is especially important for complex problem solving in an agile product development environment.

But the equation is not simply more people = more diversity, more knowledge, and more work power.

Just working in a group of people does not necessarily result in a more creative outcome than individual work. In fact, the opposite can also happen. Small group research, especially in lab setups for the idea generation and brainstorming phase, found various effects stemming from social interaction that negatively influence

A. Menning (✉) · B. Ewald · C. Nicolai · U. Weinberg
Hasso Plattner Institute for Digital Engineering, Campus Griebnitzsee,
14482 Potsdam, Germany
e-mail: axel.menning@hpi.de

B. Ewald
e-mail: benedikt.ewald@hpi.de

C. Nicolai
e-mail: claudia.nicolai@hpi.de

U. Weinberg
e-mail: uli.weinberg@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_8

creativity. Among these effects are *idea fixation* (Jansson and Smith 1991; Purcell and Gero 1996), *social loafing* (Latane et al. 1979), *evaluation apprehension* (Diehl and Stroebe 1987) and *groupthink* (Janis 1972).

The creative potential of a team only reliably surpasses the creative potential of the individual if the team interaction is consciously shaped towards it. The first step is to encourage and consider multiple perspectives. Similar to those iconic Design Thinking pictures with whiteboards heavily loaded with sticky notes, the team's mental space looks alike—very colorful, very crowded. Creating and tending to such a pluralistic batch of new ideas requires collaborative divergent thinking, as established by Guilford in his 1950 paper.

Divergent thinking consists broadly of the development of many *novel* ideas (in response to same stimulus/problem statement) and is, paradigmatically speaking, the first aspect of an outcome to be classified as "creative".[1] The most common divergent thinking exercise is any form of brainstorming. It is safe to say that brainstorming studies, i.e. studies of divergent thinking, have been among the most common studies in the field of creativity during the last 50 years. This has several, mostly pragmatic, reasons. Brainstorming studies e.g. do not need a lot of resources, especially as lab experiments are rather easy to set up and to evaluate and have a well-established framework for further analysis readily available. This framework is based on the most popular creativity test, the Torrance Test of Creative Thinking by E. Paul Torrance (TTCT 1966). The TTCT captures four important dimensions of creative thinking:

- fluency (the total amount of ideas produced)
- flexibility (the number of categories these ideas can be clustered in)
- elaboration (the level of detail of the ideas)
- originality (how often they came up in relation to all responses).

But divergent thinking as a standalone activity is not exactly creative. Even the TTCT does not fully capture Torrance's own definition of creativity,[2] as it misses out on the convergent part of creativity in both problem and solution space—nailing down the problem in the first place ("identifying the difficulty"), making the initial idea testable, testing it, and also communicating it (see Chase 1985, for further discussion).

However, sole divergent thinking leads to what has been called "pseudocreativity" (Cattell and Butcher 1968: 271) or "quasicreativity" (Cropley 1999: 89)—that is, mere novelty without any effectiveness. To achieve idea effectiveness, ideas need to be further elaborated, formulated (to achieve closure and to be communicated), evaluated and eventually validated (Cropley 2006). This is where the thinking mode of convergent thinking comes in.

[1] According to the "standard definition of creativity" by Runco and Jaeger (2012), creative ideas are characterized by their novelty and their usefulness.

[2] Torrance (1966, p. 6) defined creativity as "a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on; identifying the difficulty; searching for solutions, making guesses, or formulating hypotheses about the deficiencies: testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results."

Convergent thinking is the ability to evaluate a set of given ideas and to identify or deduce the best option (Guildford 1950; Finke et al. 1992). It relies heavily on knowledge and the ability to compare and synthesize it.[3] The endpoint of convergent thinking roughly corresponds to the second criterion for creative ideas, which is *useful* (Amabile 1983), *appropriate* (Sternberg and Lubart 1999: 3) or *valuable* (Boden 2009: 24). Creativity therefore is an interplay of divergent and convergent thinking (Finke et al. 1992). Many and different perspectives are first created through a certain flexibility of thought. Second, they are synthesized through association to get to a novel and useful concept. But how does a team get there through interaction? What are the guiding principles that distinguish an average conversation from a highly creative conversation?

Our 2018 Design Thinking Research chapter, "… and not building on that: The Relation of Low Coherence and Creativity in Design Conversations," explores the divergent side of design conversations. It presents how new ideas are generated through local disruptions (Menning et al. 2018). Accordingly, local disruptions (in form of local low coherent statements) are the linguistic equivalent of mental focus shifts, which stimulate or represent the creation of new ideas. Characteristically, good design teams pay attention to low coherent statements. They do not produce more (rather actually fewer[4]) low coherent statements, but they discuss and explore those statements more deeply and treat them as the potential missing link.

The convergent side of design conversations is represented by probing low coherent statements for potential integration into the overall discourse and making remote associations. Ideally, new contributions are combined or blended with existing pieces of knowledge.

Good ideas emerge through conversations that happen predominately in the domain of divergent flexibility (local disruption) and through convergent integration these ideas pass into the overall discourse (global integration).

2 The Coherence Style Framework (CSF)

In the following, we introduce the Coherence Style Framework (CSF) that helps to identify and analyze how divergent and convergent thinking look like on a conversational level. Topical relations are described in terms of their grammatical and lexical relation (cohesion; Halliday and Hasan 1976) and their perceived semantic connectedness (coherence; bibliographic overview in Bublitz and Augsburg 1999). The cohesion of text and talk can be objectively assessed. But to know if a contribution is off-topic and if so, how far, depends on the individual reading and sense-making of the communicative situation. This makes coherence highly subjective and hard to

[3]For a deeper, historical discussion see Cropley 2006.

[4]Goldschmidt (2014) and Suwa and Tversky (1997) found that ill-structured conversations represent ill-structured design processes, whereas good design teams converse on longer internally coherent episodes, representing some form of deep thought modus on a certain design issue.

measure (cf. Menning et al. 2017). Hence, coherence is the perceived connectedness of two discourse units. The CSF represents the coherence style of a discourse element in two dimensions: intensity and locality.

Intensity: Coherence can be described on a continuous scale between low and high. High coherence refers to a very close topical connection and a big semantic overlap of two discourse units. Low coherence describes the perceived high semantic distance between two discourse units.

Locality: Coherence can be described in two discrete states: local and global. The determination of these states depends on the size and position of two discourse units. Local coherence describes the perceived relation between subsequent and rather small discourse units. For example, the close topical relation between two subsequent sentences is described by local coherence. Global coherence refers to "the ways in which the larger segments of discourse relate to one another" (Grosz et al. 1983: 44). The global coherence definition in this text is slightly different. Global here means the semantic relation of a discourse unit to one or multiple discourse units that are not in the direct neighborhood of the discourse element in question.

Based on the distinction between intensity and locality of coherence, utterances are represented in the CSF (Fig. 1).

Each quadrant represents a certain coherence characteristic. The reading of the CSF requires a retrospective view on a conversation. Thus, the CSF can only be applied as an analytical tool after a conversation has taken place. The conversation must have a clear beginning and ending (fixed corpus size). This means that for each discourse unit at a certain time coherence information to past and future discourse units exists.

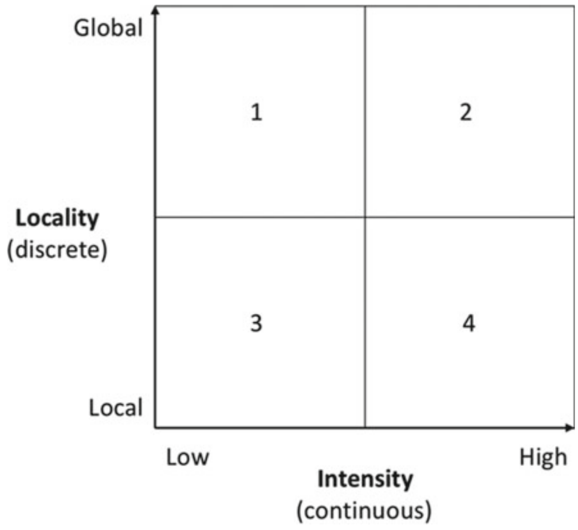


Fig. 1 The Coherence Style Framework (CSF)

Please note that the dimension of locality is discrete and the dimension of intensity is continuous. Hence the boundary between local and global is clearly defined, while the continuous scale implies that most of the coherence values are in-between extremes. This means when representing a conversation in the CSF, a threshold must be applied to determine what is “low” and “high”.

For our purpose we define speaker turns as smallest conversational units in the CSF. A speaker turn (in short: turn) is a distinct verbal contribution of a team member at a certain time. A turn begins “when a speaker begins to speak and ends when the speaker ends her or his articulation deliberately or is interrupted” (Menning et al. 2017: 2).

In the following we would like to further explain the coherence characteristics of the four different quadrants:

Global low coherence (Quadrant 1). If a turn shows global low coherent characteristics it means that it exhibits no or only weak links to what has been said before or what will be said. These turns have no explicit impact on the conversation. They do not conclude or integrate what has been said. Nor do they trigger future speaker turns to refer to it. Goldschmidt calls these turns “orphan moves” (2014).
Global high coherence (Quadrant 2). A speaker turn is globally high coherent if it shows above-average semantic similarity to speaker turns in the past (but not the preceding speaker turn) or future. These turns either conclude or integrate what has been said before or they contain information that is picked up one or multiple times in future. Global high coherent turns are similar to Goldschmidt’s critical moves (2014).
Local low coherence (Quadrant 3). A local low coherent turn shows no semantic connection to its preceding turn. Local low coherent statements often reflect mental focus shifts on the individual level. On the team level, the utterance of a local low coherent statement by one team member disrupts the thought process of all team members (Menning et al. 2017). Given this definition, we can now specify that by off-topic contributions we mean local low coherent statements. These turns are the initial elements for the idea generation sequence discussed in Sect. 4.
Local high coherence (Quadrant 4). A local high coherent turn continues the topic of its preceding turn. Discourse participants generally intend to achieve high coherence. This is known as the ‘coherence assumption’ (Graesser et al. 1994) and is a crucial element of sense-making. However, a conversation that exclusively consists of local high coherent turns is unlikely. It would resemble something between association chain exercises and small talk.

Design conversations contain low coherent statements “for good”. To a certain extent every conversation exhibits a tension between the “need for renewal and progression” (Korolija and Linell 1996: 799) and the need for sense-making, between low and high local coherent statements.

Design conversations happen to resolve ill-defined problems (Cross 2011). They contain vague language (Glock 2009) and are highly progressive by definition. Therefore, low coherent contributions have natural and frequent occurrence in design conversations. Accordingly, this chapter is not occupied with how to converse off topic (although producing good off-topic contributions is a mastery itself), but rather how to deal with off-topic contributions productively.

Designers probe the potential of local low coherent turns and based on the CSF, we can capture this probing activity.

3 The Interplay Between Local Disruption and Global Integration

We can now assess which possible transitions between local disruption and global integration exist. For example, if perceiving a local disruption, it would be possible to follow up with another low coherent turn (global low coherence or local low coherence). The sequence of two subsequent topical disruptions often comes up when the initial off-topic contribution is ignored and the next person jumps back to the previous topic. Another situation in which two subsequent off-topic contributions may occur would be to "fire back" by responding to a verbal disruption with another verbal disruption. In both cases, the creative potential of the initial off-topic contribution is not further explored. A chain of multiple subsequent disruptions can also be observed in brainstorming when a list of ideas is generated. Within this list, while one idea may not necessarily pick up the topic of the idea before, they all relate to the topic of the list, which is the global topic of the design issue. In this situation, local disruption and global coherence exist simultaneously. This example shows that the sequencing of coherence styles has an extension: two coherence styles may exist at the same time.

The activity of building lists of ideas (also known as ideation or brainstorming) qualifies for the first criterion of idea generation (having many different new ideas), but it does not necessarily secure the second criterion, which is about the usefulness of ideas. Conventional brainstorming is efficient, because it promotes free and associative thinking. These brainstorming techniques leave the assessment of the creative potential of an utterance implicit and, in the interest of creating many and new ideas, do not make it a collective matter (this is usually suggested by the prompt "defer judgement").

The brainstorming technique *silent brainstroming* is a classic example for list creation. When performing *silent brainstroming*, the team members are invited to create as many ideas as possible for a certain amount of time. The ideas are not shared immediately but shared afterwards. This technique is highly effective in terms of idea quantity, but it does not necessarily make use of the full potential of the group. For *silent brainstorming* the argument more people = more ideas may hold true (better overall fluency), but this does not automatically mean more people = different (better) ideas, e.g. better cumulated flexibility, novelty or degree of elaboration.

When selecting brainstorming techniques, it is crucial to find a good balance between a mere quantity of ideas and amount of collective idea generation. In other words: The chances of having a good idea by having many ideas counter the chances of developing a good idea out of any idea.

We have mentioned that brainstorming as list creation is reflected in many local low coherent turns that are at the same time globally coherent.

Given that a team works well (in terms of team dynamics, team cohesion and psychological safety) there are alternative routes of idea generation, not via lists and close to what a natural discussion is. Instead of creating lists, the potential of a low coherent input is examined collectively. This eventually leads to a meaningful integration into the design issue at hand or the overall discourse. Representing this procedure in the CSF would mean that local low coherent statements are proceeded with a sequence of local high coherent statements (exploring the local low coherent statement). These statements are again proceeded by a global high coherence statement signaling the integration of the off-topic contribution into the general design conversation (see Fig. 2).

This procedure makes up the title of this book chapter: Collective creativity is an interplay between local low coherence and global high coherence.

Off-topic turns invite participants to infer connection (cf. Grosz et al. 1995). This is relevant for idea generation. Being exposed to and exploring the meaning of low coherent statements increases the likelihood of creating new ideas. Off-topic contributions shift the team members' focus of attention. These focus shifts are beneficial for idea generation (Suwa and Tversky 1997). This is especially the case if the creative value of the initial disruptive contribution cannot be directly realized, but in search of a meaningful connection other, new associations and ideas come to mind. Similar principles of idea generation are reported by Einstein, (cf. combinatorial play, 1954), Koestler (cf. Bisociation, 1964), and Finke et al. (cf. conceptual blends, 1992). Ideally, these local low coherent turns are integrated back into the greater picture of

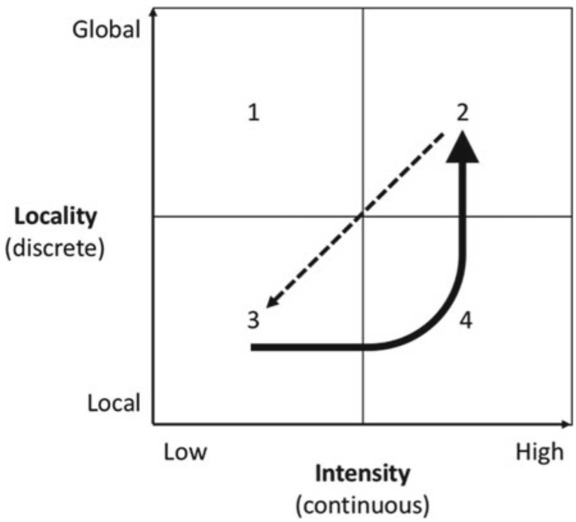


Fig. 2 Collective exploration of low coherent turns as an alternative to conventional brainstorming

the design. Therefore, once the idea of an off-topic contribution is examined and turned into a manageable proposition, the next step is to explore ways to re-integrate it. Either one is able to relate the proposition to a specific statement that has been brought up earlier, or it is integrated into a bigger discourse segment topic (e.g. design issue). This happens rather explicitly by proposing what the relationship could be and in which way the off-topic contribution makes sense. The sequence of creative topic treatment ends in global high coherence. It can be reinitiated as soon as a perceived low coherent statement disrupts the course of the conversation again. The dashed line in Fig. 2 proposes that the sequence has cyclic characteristics. This means ideally the team cycles through the quadrants over and over again. Different existing ideation techniques work that way. Take, for example, the brainstorming method "What would XYZ do?", where XYZ is substituted with a well-known and distinguished character such as Superman. The well-known attributes of that person are then used to approach a certain issue with this new perspective. Or the ideation method called *idea blossoming*: The team members are asked to pick another idea of a team member and then build around this idea eight more that are further elaborating the critical functions and features of the initial idea concept. Many variations of these ideation techniques exist. They work on the principles of associative and lateral thinking (de Bono 1991; Mednick 1962; Sternberg and Lubart 1993), and they have proven to produce a lateral variety of ideas. Of course, the effectiveness of these techniques depends on the experience of the team applying them. In conclusion: Idea generation can be most efficiently facilitated if different brainstorming techniques (list creation and collective exploration) are combined and balanced.

4 Implications

There is a lack of exercises that systematically advance team-skills in merging, combining and integrating a batch of diverse pieces. We have shown that it is crucial to design conversations to work with off-topic contributions and to put them forward. We therefore need to build more systematic training exercises for high-quality idea generation in teams.

Of course, the direct examination of ideas limits the number of ideas to be produced. It is about finding a good balance of having enough low coherent input, and actually working with it. With this book chapter we want to make practitioners aware that the ratio of quantity of ideas and instant examination of an idea is a factor to play with.

Some more general implications can be drawn. On a theoretical level, this work brings us closer to understanding the complexities of collective creativity. Most ideas in design thinking are not centrally planned and not the work of a single creator. Instead, novel and useful ideas arise out of the interplay of divergent flexibility and convergent combination/synthesis and thorough validation.

On a practical level, this paper underpins the relevance of teamwork. Based on the CSF, we call for more training and methods that build up awareness for exploring and integrating local low coherent statements into "the bigger picture". Once a team feels proficient in interacting local low coherent and global high coherent at the same time, they will always have a standardized rhythm to rely on, which can also prevent them from getting stuck (cf. design fixation).

We primarily address design conversations. But we assume that this model is of such robustness that it can be extended to other interactional activities within the creative domain.

References

Amabile, T. M. (1983). *The social psychology of creativity*. New York: Springer.

Boden, M. A. (2009). Computer models of creativity. *Interfaces, 1993*(April), 80–83.

Bublitz, W., & Augsburg, U. (1999). A bibliography of coherence and cohesion. In W. Bublitz, U. Lenk, & E. Ventola (Eds.), Coherence in spoken and written discourse. How to create it and how to describe it (pp. 267–295). Amsterdam: John Benjamins.

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

Chase, C. I. (1985). Review of the torrance tests of creative thinking. In J. V. Mitchell Jr. (Ed.), *The ninth mental measurements yearbook* (pp. 1631–1632). Lincoln: University of Nebraska, Buros Institute of Mental Measurements.

Cropley, A. (1999). Definitions of creativity. In M. Runco & S. Pritzker (Eds.), *Encyclopedia of creativity* (pp. 511–524). San Diego: Academic Press.

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

Cross, N. (2011). *Design thinking: Understanding how designers think and work*. Oxford: Berg Publishers.

de Bono, E. (1991). Lateral and vertical thinking. In J. Henry (Ed.), *Creative management* (pp. 16–23). London: Sage.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53,* 497–509.

Einstein, A. (1954). *Ideas and opinions*. New York: Crown.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and Applications*. MIT Press.

Gilson, L. L., Lim, H. S., Litchfield, R. C., & Gilson, P. W. (2015). Creativity in teams: A key building block for innovation and entrepreneurship. In C. E. Shalley, M. A. Hitt, & J. Zhou (Eds.), *Oxford library of psychology. The Oxford handbook of creativity, innovation, and entrepreneurship* (pp. 177–204). New York, NY, USA: Oxford University Press.

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

Goldschmidt, G. (2014). *Linkography: Unfolding the design process*. Cambridge, MA: MIT Press.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–395.

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1983). Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting on Association for Computational Linguistics* (pp. 44–50).

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics, 21*(2), 203–225.

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

Halliday, M. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Boston: Houghton Mifflin.

Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies, 12,* 3–11.
Koestler, A. (1964). *The act of creation*. New York, NY: Macmillan.
Korolija, N., & Linell, P. (1996). Episodes: Coding and analyzing coherence in multiparty conversation. *Linguistics, 34*(4), 799–831.
Latane, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37,* 822–832.
Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69,* 220–232.
Menning, A., Ewald, B., Nicolai, C., & Weinberg, U. (2018). "… and not building on that": The relation of low coherence and creativity in design conversations. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research. Understanding innovation*. Cham: Springer.
Menning, A., Grasnick, B. M., Ewald, B., Dobrigkeit, F., & Nicolai, C. (2017). Combining computational and human analysis to study low coherence in design conversations. In B. T. Christensen, L. J. Ball, & K. Halskov (Eds.), *Analyzing design thinking: Studies of cross-cultural co-creation*. Leiden: CRC Press/Taylor & Francis.
Purcell, A. T., & Gero, J. S. (1996). Design and other types of fixation. *Design Studies, 17*(4), 363–383.
Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092.
Sternberg, R. J., & Lubart, T. I. (1993). Investing in creativity. *Psychological Inquiry, 4,* 229–232.
Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 3–15). Cambridge: Cambridge University Press.
Suwa, M., & Tversky, B. (1997). What do architects and students perceive in their design sketches? A protocol analysis. *Design Studies, 18*(4), 385–403.
Torrance, E. P. (1966). *The torrance tests of creative thinking–Norms: Technical manual research edition—Verbal tests, forms A and B—Figural tests, forms A and B*. Princeton, NJ: Personnel Press.
Valkenburg, R. (2000). *The reflective practice in product design teams*. Delft: Delft University of Technology.

The Neuroscience of Team Collaboration During a Design Thinking Event in Naturalistic Settings

Naama Mayseless, Grace Hawthorne and Allan Reiss

Abstract While previous studies have begun to investigate the neuroscience of design thinking, current knowledge is still limited. One limitation faced to date, is the need for a naturalistic design methodology that can incorporate the inherit properties of design thinking in a neuroscientific design. Here we will introduce the concept of neuroscience in design thinking both at the individual level as well as in teams and propose an experimental design to study the neuroscience of design that imposes as little constraints as possible on the natural flow of team collaboration during a design thinking event.

1 Introduction

Design thinking has been investigated extensively over the last decade with research moving from a view of design thinking as a set of cognitive abilities that can lead to better problem solving, to a more subjective and intuitive view of design thinking (Simon 1973; Dinar et al. 2015; Lazar 2018). Neuroscience has the ability to build upon the evolution of design thinking by uncovering the underlying neural processes that occur during individual or group/team-based design thinking. Identifying the neural processes occurring during an innovation event opens up ways to potentially increase the impact of the event. Design tasks are often termed "ill-structured" problems (Simon 1973) or "wicked problems" (Buchanan 1992), where the problem is

N. Mayseless (✉) · A. Reiss
Center for Interdisciplinary Brain Sciences Research, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, 401 Quarry Road, Stanford, CA 94305-5795, USA
e-mail: nmay@stanford.edu

A. Reiss
e-mail: areiss1@stanford.edu

G. Hawthorne
Hasso Plattner Institute of Design (d.School), Building 550, 416 Escondido Mall, Stanford, CA 94305-3086, USA
e-mail: grace@dschool.stanford.edu

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_9

not well-defined, and has an open-ended flow without an obvious solution. These are intentional characteristics of typical design thinking problems. Therefore, the challenge is to develop experimental settings that allow examination of the interrelations between brain activity during a design thinking event, especially those in more naturalistic settings that take into account the open-endedness of the challenge or problem.

The first part of this chapter will introduce the concept of neuroscience in design thinking, both at the individual level as well as in teams. In the second part of the chapter we propose an experimental design to study the neuroscience of design that imposes as little constraint as possible on the natural flow of a collaborative team during a design thinking event.

1.1 Neuroscience of Design Thinking

Many methods used in neuroscience have inherent limitations that can affect the effectiveness of naturalistic type studies. These include being susceptible to movement artifacts in an MRI scanner, noise from external sources in an EEG examination, and the need to be in a highly structured laboratory setting for many other approaches. In particular, most neuroimaging methods do not lend themselves easily to more complex social interactions, such as those required in everyday design team interactions. Despite these limitations, several studies have focused on the neuroscience of design thinking, either directly or through the lens of studying problem solving and creativity. For example, using functional magnetic resonance imaging (fMRI), Kowatari et al. (2009) studied the effect of design training novices and experts who were asked to design a new pen while being scanned with fMRI. Results suggest that experts tend to recruit the right prefrontal (PFC) and parietal cortices to a greater extent than novices who used both the right and left equally. These results indicate that design experts may have a more efficient profile of brain activation during a design thinking task compared to novices. In addition, this study reported that the originality of the designs was related to the interaction between right and left brain activity but not directly to either one. These results indicate that training in design has an effect on the topography of brain activation. In another study, Alexiou et al. (2009) asked participants to design a room given the instruction that the room be functional, comfortable and contain at least a bed, a wardrobe and a desk. The investigators compared this task to a control task where the exact location of the furniture was laid out in the instructions. Results showed increased activation in the right dorsolateral prefrontal cortex for design compared to the control task (Alexiou et al. 2009; Gilbert et al. 2010).

These studies suggest that the prefrontal cortex is involved in design thinking. The PFC has connections with brain regions associated with motor control, performance monitoring and higher order sensory processing (association cortex and parietal cortex). The PFC has broadly been seen to be activated in tasks requiring executive control, such as inhibition and switching between modes of thinking (Garavan et al.

1999; Aron et al. 2004). In addition, extensive evidence from the neuroscience of creativity suggest that the PFC is important for creative ability (Limb and Braun 2008; Beaty et al. 2015; Mayseless and Shamay-Tsoory 2015; Saggar et al. 2015).

In another interesting study investigating the neuroscience of design creativity, investigators asked participants to design a book cover based on a book description and then to evaluate their ideas, while they were scanned with fMRI (Ellamil et al. 2012). The results indicated that generating ideas preferential recruited the medial temporal lobe (including the hippocampus and parahippocampus), while evaluating ideas was associated with increased activity in executive and default network regions (including prefrontal regions, inferior parietal lobule and temporal regions). This study suggests that it may be the recruitment of evaluation processes that account for observed activation in the PFC during a design task, and demonstrates the importance of investigating the underlying stages of the design process.

The aforementioned studies of design thinking, although few in number, are important advancements in the study of the neuroscience of design thinking. Building on the extant studies available to date, in a recent review of the cognitive neuroscience of design creativity Lazar (2018) concluded that for design thinking to be better understood, study designs using real-life design tasks should be employed.

1.2 Neuroscience of Design Teams

Another important aspect of design is the ability to work in teams as part of the design process. Collaboration is especially important as it is often assumed that groups of individuals can work together to solve complex problems they are unable to solve on their own. While creativity in the design process has traditionally been regarded and researched as an individual trait, there is increasing interest in the ability of groups to design and create innovative ideas and products (Baruah and Paulus 2009). Previous research has begun to examine the process by which problem solving occurs in teams, as well as the different types of collaboration that can occur during different phases of design thinking (Stempfle and Badke-Schaub 2002). Despite this progress, a clear, brain-based model that informs how team interactivity contributes and impacts the outcome of an innovation event is lacking.

1.3 Hyperscanning as a Promising Measure of Social Interaction

Recently, a novel imaging technique that allows for simultaneous measurement of brain activation from more than one individual was developed and termed "hyperscanning" (Montague et al. 2002; Cui et al. 2012). Since the introduction of this method, many hyperscanning studies have been performed and published, adding to our understand of social interaction. Hyperscanning allows us to measure the degree of brain-to-brain synchronization during a social interaction (Fig. 1). Brain-to-brain synchrony is quantified as the amount of correlation or coherence in activation between two individuals and is often termed inter-brain synchrony (IBS). For example, it has been found that IBS level correlates with the level of comprehension between partners (Stephens et al. 2010) and that IBS increases during cooperation but not during competition (Cui et al. 2012). In addition, the pattern of IBS increase due to cooperation depends on the gender diversity of the team (Baker et al. 2016).

Despite this new advancement in neuroimaging allowing for more real-life approximation of social interactions, tasks used are still far from mimicking real-life. For example, tasks involving conversation often require participants to take turns every set amount of time (Hirsch et al. 2018), take turns singing parts of a song (Pan et al. 2018), or take turns generating ideas in a creativity task (Lu et al. 2018; Xue et al. 2018). The use of repeated trials as part of a methodological neuroimaging study may interrupt the natural flow of social interaction, challenging our ability to generalize the results to real life scenarios. To our knowledge only one group of researchers has published work using a continuous verbal communication paradigm, without imposing turn-taking (Jiang et al. 2012; Jiang et al. 2015). These researchers have focused their work on the distinct pattern of face-to-face communication.

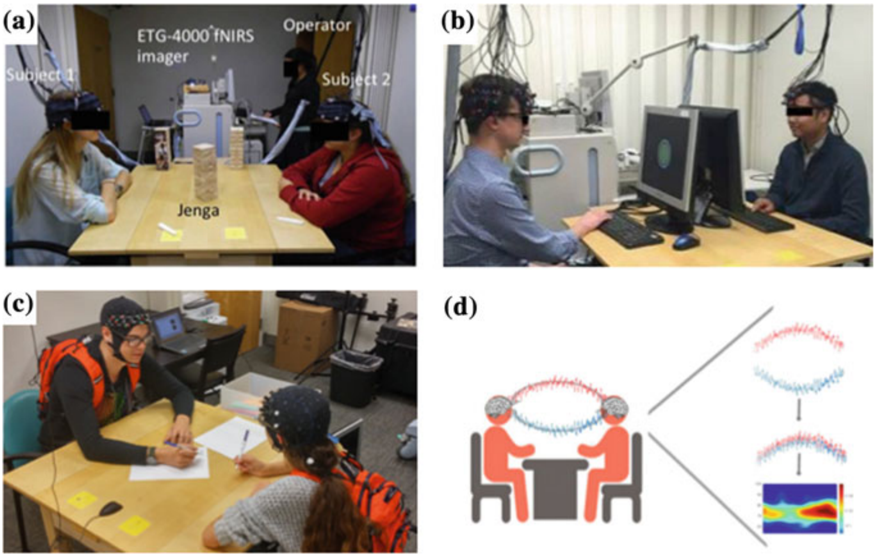


Fig. 1 functional NIRS hyperscanning. **a** experimental setup reproduced from Liu et al. (2016) of two participants cooperating while playing the game of Jenga™. **b** Experimental setup of cooperation reproduced from Baker et al. (2016). **c** Experimental setup of the current study of design creativity. **d** Example of inter-brain synchrony analysis: on the left two people interacting; on the right NIRS signals from partner 1 and partner 2 are combined and checked for synchrony

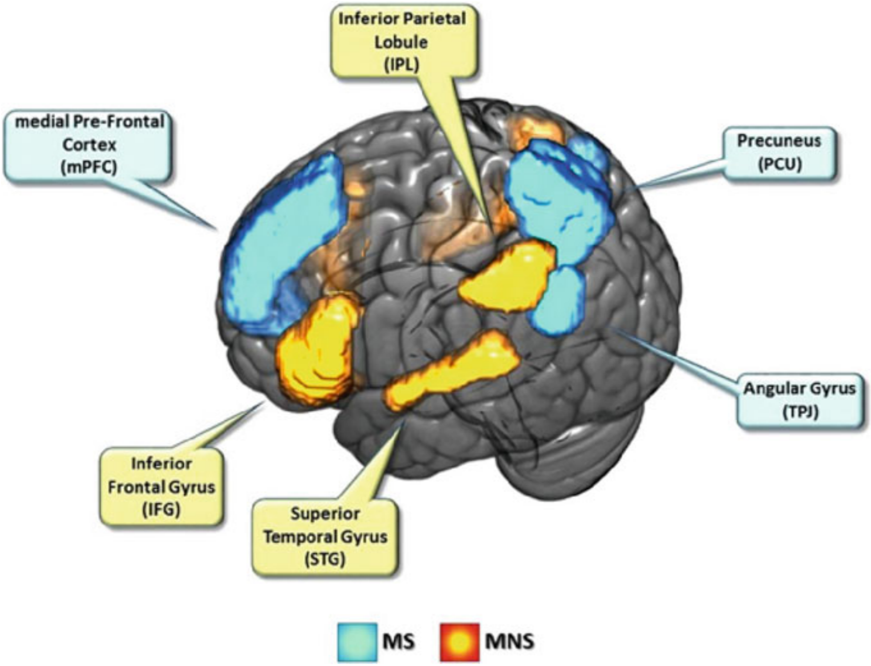


Fig. 2 Two main brain systems involved in social interaction. This pictures was adapted from Begliomini et al. (2017) under the Creative Commons Attribution 4.0 international license

Despite these limitations, there is a great deal to be learned from studies involving IBS measurements. A recent review of hyperscanning during social interaction (Wang et al. 2018) found two main neural networks often involved in IBS (Fig. 2). One is the mirror neuron system (MNS; including the inferior frontal gyrus (IFG), inferior parietal lobule (IPL), and the superior temporal gyrus (STG)), while the other is the mentalizing system (MS; consisting of the temporal-parietal junction (TPJ), precuneus and prefrontal cortex). The MNS is generally involved in imitation or even observing others' actions and movements, and is evident both in animals as well as in humans (Iacoboni and Dapretto 2006). The MS is often seen in relation to trying to understand others' intentions and emotions based on gestures, behaviors and facial expressions (Frith and Frith 2006).

2 Experimental Design

Our purpose was to achieve a naturalistic design thinking session, while imposing as little constraint as possible on the process. Compared to previous studies that have used timed turn taking paradigms, we adopted a natural conversation flow similar to what would occur in a design thinking session outside of the lab setting.

The study design presented here was developed to investigate neural synchrony using functional near-infrared spectroscopy (fNIRS) in a naturalistic, interactive setting during which subjects are able to orally communicate face-to-face while solving a real world problem. fNIRS is a promising method for investigating the brain basis of interpersonal interactions in natural settings. Functional NIRS is a robust, non-invasive optical imaging method that measures changes in cerebral blood flow in a manner similar to fMRI. Functional NIRS also has the advantage of being portable, low cost and less prone to movement artifact than other imaging methods (Monden et al. 2012). In addition, fNIRS measures changes in both oxy- and deoxy-hemoglobin and provides higher temporal resolution than fMRI (Cui et al. 2011), yet with better spatial resolution than electroencephalography (EEG). It is especially useful for paradigms that require a more natural setting. Indeed, an increasing number of fNIRS hyperscanning studies have been recently published. These studies investigate INS during interactions, both verbal (Jiang et al. 2012; Hirsch et al. 2018), semi-verbal [cooperative singing/humming (Osaka et al. 2015)], and nonverbal types (Funane et al. 2011; Cui et al. 2012; Holper et al. 2012).

In our study, participants were randomly assigned to two-person teams to collaboratively solve a problem, either requiring creativity (product design) or following explicit instructions (model building). Our goal was to utilize a design that would achieve as close to a real-world environment as possible. Therefore, dyads were asked to work on problems for a continuous time of 10 min with little instruction and no interventions. All the dyads were instructed to solve the same problem and build the same 3D model. Changes in cerebral activity during the tasks were continually recorded using a fNIRS-based system.

In addition to the fNIRS session, we assessed individual divergent thinking abilities, creative achievement and collaboration indices.

2.1 *Experimental Procedure and Tasks*

Participants in each dyad were seated in-front of one another on opposite sides of a square table (Fig. 3). The experimental procedure consisted of one 10-minute creative design thinking session and one 10-minute control 3D model building session. The creative design and 3D model building sessions were counterbalanced. During the creative design session, dyads were asked to work together to design a product that would motivate people to vote. The product could be of any design or material. Dyads were also told that they would be asked to explain their product to an investigator after the completion of the session. During the control 3D model building session, dyads were asked to work together to build a model of an airplane. They were presented with instructions for completing the 3D model (Fig. 3) and given 10 min. The control task was chosen in order to control for creative design while still requiring teams to collaborate.

Fig. 3 Experimental setup of the current study of design creativity

2.2 *Post Experiment Assessments*

After the NIRS session, each participant completed an additional assessment session in a separate room during which they completed tasks measuring creative divergent thinking, executive functions and general intelligence. After completion of the assessments, participants were sent links to additional surveys to be completed online in their free time. Surveys included demographic information, Creative Achievement Questionnaire [CAQ; (Carson et al. 2005)] and Revised NEO-FFI Personality Inventory (NEO PI-R) that examines a person's Big Five personality traits (openness to experience, conscientiousness, extraversion, agreeableness, and neuroticism) (Costa and McCrae 2008). The CAQ is a reliable and valid measure of creative productivity across ten domains including visual arts, music, creative writing, dance, drama, architecture, humor, scientific discovery, invention, and culinary arts.

2.3 *Creative Divergent Thinking (DT)*

Participants completed the AUT and figural subset of the TTCT, The Torrance Test of Creative Thinking, Figural (TTCT-F) is a paper–pencil based valid and reliable assessment measure of the divergent thinking aspects of creativity (Torrance 1974). The measure takes approximately 30 min to complete and has two equivalent versions (A and B). We administered all three subtests of the TTCT-Figural assessment: Picture Construction, Picture Completion, and Parallel Lines. For each task, participants produce as many figural drawings as possible in a pre-determined period of time (∼10 min). For example, stimuli include parallel lines in boxes on the page with instructions that direct participants to create as many different and novel pictures with the parallel lines that they can in 10 min. Standard scores for fluency and originality were used as measures. TTCT-F tests were scored by Scholastics Testing Service, Inc (http://ststesting.com/). Raters were not aware of the experimental design.

2.4 Alternate Uses Task (AUT)

The AUT task is a well-known, reliable and valid task that examines DT (Guilford 1967). In the present study participants were presented with a printed list of five common objects and asked to list as many alternative uses as possible for each object, within a time limit of 10 min (total time for all objects). The items were a cardboard box, car tire, pencil, paper clip and drinking glass. The most common everyday use was indicated in parenthesis next to the name of each object. Before performing the task, participants were presented with an example of uses for a newspaper in order to familiarize them with the task. Only responses that did not replicate the common uses given were counted and included. Scoring included fluency (number of responses), and originality (rarity of the response). Final scores were calculated based on the average score of all items. Scoring of originality followed Torrance (1974). Original responses were defined as statistically infrequent responses within the population of the study.

2.5 Executive Functions

2.5.1 The Delis–Kaplan Executive Function System (D-KEFS)

The Color-Word Interference Test (CWIT) subset of the D-KEFS was used as an executive functions assessment (Delis 2001). The CWIT is based on the Stroop measure (Stroop 1935) and it consists of four conditions. The first two conditions (color-naming and word-reading) assess processing speed, whereas the last two assess “higher-level” inhibition and cognitive flexibility.

2.6 The Wechsler Abbreviated Scale of Intelligence-II

The Wechsler Abbreviated Scale of Intelligence-II was used to measure general intelligence (WASI-II, Wechsler 2011). The WASI-II is designed to be administered individually in approximately 30 min. The measure consists of 4 subtests: Vocabulary, Similarities, Block Design, and Matrix reasoning used to obtain Full Scale IQ (FSIQ). The WASI-II has a mean of 100 and a SD of 15.

2.7 Task Related Assessments

2.7.1 Subjective Collaboration Index

Participants were asked to rate the overall collaboration of the team, the collaboration rating of themselves and their partner and the success of the session, on a scale of 1 (least) to 5 (most). They were also asked about the final product.

2.7.2 Coding of Collaboration and Leadership

Two trained raters independently assessed collaboration and leadership of each dyad during the NIRS session. Each 10-minute session (creative design and control condition) were divided into 2-minute segments to allow for better temporal granularity of the team process while still allowing for the time-consuming process of coding behavior from video. Raters watched videos of the session and scored each 2-minute segment on the degree of collaboration in each dyad on a 7-point Likert scale from 1 (poor collaboration) to 7 (very good collaboration). Inter-rater reliability index (as measured by Intra Class Correlation Coefficient (ICC)) was satisfactory for all measures (collaboration design ICC:0.88, model ICC:0.77). Raters also rated the degree of leadership from 0 (equal contribution) to 7 (dominance of conversation and ideas) with good agreement (ICC: 0.63).

2.7.3 Coding of Creative Design Ideas

Two trained raters independently assessed the quality of product ideas on a 5-point Likert scale from 1 (low) to 7 (high). Raters viewed all videos before scoring. Product ideas were scored for originality (how original and infrequent the idea is) and efficacy (how efficient the product is, whether it improved/incentivized time/cost/accessibility of voting). Inter-rater reliability index (as measured by Intra Class Correlation Coefficient (ICC)) was satisfactory for all measures (originality: ICC = 0.78, efficacy: ICC = 0.86).

2.8 Assessing IBS of Design Teams

In order to investigate the IBS model of creative design teams, continuous wave fNIRS (tandem NIRSport; LLC NIRx Medical Technologies) was used. We calculated the IBS between individuals assigned to the same team and compared it to the IBS of permuted teams—i.e. individuals assigned to different teams. In addition, to tease apart the element of creative collaboration, we compared IBS between the creative design task and the 3D model task.

To further investigate the dynamical progression of IBS and its relation to collaboration, we divided the 10-min task into 2 min segments that were then averaged to obtain 5 consecutive averaged time points, similar to segmentation utilized for behavioral coding of the session videos.

Finally, we assessed for IBS-behavior correlations that might help explain the experimental data and connect the IBS data to observed behavior and design output.

3 Implications and Future Activities

By using fNIRS imaging data, including the dynamical changes in IBS across the design task, and a naturalistic study design to investigate areas of the brain that map to team collaboration and assessing collaborative outcomes, we will study the impact and value of team collaboration. This pioneering project is designed to understand the impact of team collaboration during the innovation process. Our overarching goal is to uncover new information that will improve team-based problem-solving in the 21st century. In this unique study of team-based design and innovation, we will utilize measures of inter-brain synchrony within teams to explicate the brain (and correlated behavioral) factors underlying the team.

References

Alexiou, K., et al. (2009). Exploring the neurological basis of design cognition using brain imaging: some preliminary results. *Design Studies, 30*(6), 623–647.

Aron, A. R., et al. (2004). Inhibition and the right inferior frontal cortex. *Trends in cognitive sciences, 8*(4), 170–177.

Baker, J. M., et al. (2016). Sex differences in neural and behavioral signatures of cooperation revealed by fNIRS hyperscanning. *Scientific reports, 6,* 26492.

Baruah, J., & Paulus, P. B. (2009). Enhancing group creativity: The search for synergy. In *Creativity in groups* (pp. 29–56), Emerald Group Publishing Limited.

Beaty, R. E., et al. (2015). Default and executive network coupling supports creative idea production. *Scientific reports, 5,* 10964.

Begliomini, C., Cavallo, A., Manera, V., Becchio, C., Stramare, R., Miotto, D., & Castiello, U. (2017). Potential for social involvement modulates activity within the mirror and the mentalizing systems. *Scientific reports, 7*(1), 14967.

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

Carson, S. H., et al. (2005). Reliability, validity, and factor structure of the creative achievement questionnaire. *Creativity Research Journal, 17*(1), 37–50.

Costa, P. T., & McCrae, R. R. (2008). The revised neo personality inventory (neo-pi-r). *The SAGE handbook of personality theory and assessment, 2*(2), 179–198.

Cui, X., et al. (2011). A quantitative comparison of NIRS and fMRI across multiple cognitive tasks. *Neuroimage, 54*(4), 2808–2821.

Cui, X., et al. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage, 59*(3), 2430–2437.

Delis, D. (2001). *Delis-Kaplan executive function scale (D-KEFS)*. San Antonio: The Psychological Corporation.

Dinar, M., et al. (2015). Empirical studies of designer thinking: past, present, and future. *Journal of Mechanical Design, 137*(2), 021101.

Ellamil, M., et al. (2012). Evaluative and generative modes of thought during the creative process. *Neuroimage, 59*(2), 1783–1794.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531–534.

Funane, T., et al. (2011). Synchronous activity of two people's prefrontal cortices during a cooperative task measured by simultaneous near-infrared spectroscopy. *Journal of Biomedical Optics, 16*(7), 077011.

Garavan, H., et al. (1999). Right hemispheric dominance of inhibitory control: an event-related functional MRI study. *Proceedings of the National Academy of Sciences, 96*(14), 8301–8306.

Gilbert, S. J., et al. (2010). Involvement of right dorsolateral prefrontal cortex in ill-structured design cognition: an fMRI study. *Brain Research, 1312,* 79–88.

Guilford, J. P. (1967). The nature of human intelligence.

Hirsch, J., et al. (2018). A cross-brain neural mechanism for human-to-human verbal communication. *Social cognitive and affective neuroscience, 13*(9), 907–920.

Holper, L., et al. (2012). Between-brain connectivity during imitation measured by fNIRS. *Neuroimage, 63*(1), 212–222.

Iacoboni, M., & Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience, 7*(12), 942.

Jiang, J., et al. (2015). Leader emergence through interpersonal neural synchronization. In *Proceedings of the National Academy of Sciences*: 201422930.

Jiang, J., et al. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience, 32*(45), 16064–16069.

Kowatari, Y., et al. (2009). Neural networks involved in artistic creativity. *Human Brain Mapping, 30*(5), 1678–1690.

Lazar, L. (2018). The cognitive neuroscience of design creativity. *Journal of Experimental Neuroscience, 12,* 1179069518809664.

Limb, C. J., & Braun, A. R. (2008). Neural substrates of spontaneous musical performance: An fMRI study of jazz improvisation. *PLoS ONE, 3*(2), e1679.

Liu, N., Mok, C., Witt, E. E., Pradhan, A. H., Chen, J. E., & Reiss, A. L. (2016). NIRS-based hyperscanning reveals inter-brain neural synchronization during cooperative Jenga game with face-to-face communication. *Frontiers in human neuroscience, 10*, 82.

Lu, K., et al. (2018). Cooperation makes a group be more creative. Cerebral Cortex.

Mayseless, N., & Shamay-Tsoory, S. (2015). Enhancing verbal creativity: modulating creativity by altering the balance between right and left inferior frontal gyrus with tDCS. *Neuroscience, 291,* 167–176.

Monden, Y., et al. (2012). Clinically-oriented monitoring of acute effects of methylphenidate on cerebral hemodynamics in ADHD children using fNIRS. *Clinical Neurophysiology, 123*(6), 1147–1157.

Montague, P. R., et al. (2002). Hyperscanning: simultaneous fMRI during linked social interactions, Academic Press.

Osaka, N., et al. (2015). How two brains make one synchronized mind in the inferior frontal cortex: fNIRS-based hyperscanning during cooperative singing. *Frontiers in psychology, 6,* 1811.

Pan, Y., et al. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *Neuroimage, 183,* 280–290.

Saggar, M., et al. (2015). Pictionary-based fMRI paradigm to study the neural correlates of spontaneous improvisation and figural creativity. *Scientific reports, 5,* 10894.

Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence, 4*(3–4), 181–201.

Stempfle, J., & Badke-Schaub, P. (2002). Thinking in design teams-an analysis of team communication. *Design Studies, 23*(5), 473–496.

Stephens, G. J., et al. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences, 107*(32), 14425–14430.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643.

Torrance, E. (1974). *The torrance tests of creative thinking-TTCT manual and scoring guide: Verbal test a, figural test*. Lexington, KY: Ginn.

Wang, M.-Y., et al. (2018). Concurrent mapping of brain activation from multiple subjects during social interaction by hyperscanning: a mini-review. *Quantitative imaging in medicine and surgery, 8*(8), 819.

Wechsler, D. (2011). WASI-II: Wechsler abbreviated scale of intelligence. *PsychCorp*.

Xue, H., et al. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *Neuroimage, 172,* 527–537.

Mining the Role of Design Reflection and Associated Brain Dynamics in Creativity

Neeraj Sonalkar, Sahar Jahanikia, Hua Xie, Caleb Geniesse, Rafi Ayub, Roger Beaty and Manish Saggar

Abstract Reflection—the activity of reasoning through an action that has occurred—has been shown to be of importance to the development of design expertise. Although design reflection has been widely studied previously, several gaps in the knowledge still exist. First, previous work in design reflection has been mostly limited to *descriptive* and *prescriptive* research, while very few researchers investigated the effect of design reflection on performance of individuals and teams. Second, previous researchers limited the study of reflection to the language used and its reference to the design problem or solution space. Third, previous work on design reflection has not taken into account the antagonist of reflection—i.e., *rumination*. Rumination is characterized as repetitive and persistent evaluation of the meaning, causes, and consequences of one's affective state and personal concerns, and has been shown to negatively affect creativity and problem solving. In this project, we planned to address these limitations by (1) assessing the effects of different types of reflection on creative performance; (2) going beyond the frontier of language (or speech) and

N. Sonalkar (✉)
Center for Design Research, Stanford University, Building 560,
424 Panama Mall, Stanford, CA 94305-2232, USA
e-mail: sonalkar@stanford.edu

S. Jahanikia · C. Geniesse · R. Ayub · M. Saggar
Stanford University School of Medicine, St 1356, 401 Quarry Road, Stanford, CA 94305, USA
e-mail: saharjkobfuscate@stanford.edu

C. Geniesse
e-mail: geniesseobfuscate@stanford.edu

M. Saggar
e-mail: saggar@stanford.edu

H. Xie
OHSU Center for Regenerative Medicine, 3181 SW Sam Jackson Park Road,
Portland, OR 97239, USA
e-mail: xieh@ohsu.edu

R. Beaty
Department of Psychology, The Pennsylvania State University,
140 Moore Building, University Park, PA 16802, USA
e-mail: rub736@psu.edu

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_10

additionally investigating the role of brain and interaction dynamics during design reflection; and (3) including psychological construct of rumination in addition to reflection. We hypothesized that given the critical importance of reflection in design thinking, our approach will provide a comprehensive understanding of the *interplay* between brain dynamics, design reflection, and creativity.

1 Introduction

Reflection—the activity of reasoning through an action that has occurred—has been shown to be of importance to the development of design expertise. Schön (1983) put forward the theory of reflective practice to explain what designers do, beyond what is captured in the technical rationality of process models and flow diagrams. Dorst and Reymen (2004) put forward seven different levels of design expertise—beginner, advanced beginner, competent, proficient, expert, master and visionary—and posit that *reflection is critical to the development of higher levels of design expertise*. In this project, we are employing dynamical information from three different domains—i.e., brain, design interaction and speech—*to quantify the reflection process and link it to individual differences in creativity and design thinking.*

Despite being widely studied previously, our understanding of design reflection is still limited by several gaps. First, there is a lack of investigation of the effect of design reflection on performance of individuals and teams. Second, previous researchers have mainly focused on the language used and its reference to the design problem or solution space. Third, previous work on design reflection has largely ignored the nemesis of reflection—*rumination*. Rumination is characterized as repetitive and persistent evaluation of the meaning, causes, and consequences of one's affective state and personal concerns (Whiteman and Mangels 2016), which has been shown to negatively affect creativity and problem solving (Verhaeghen et al. 2005).

In this project, we aimed at addressing these limitations by (1) assessing the effects of different types of reflection on creative performance; (2) going beyond the frontier of language (or speech) to study the role of brain and team-interaction dynamics during reflection; and (3) including psychological construct of rumination. We hypothesized that given the critical importance of reflection in design thinking, our approach may shed light on the *interplay* between brain dynamics, design reflection, and creativity.

This chapter is organized as follows: in Sect. 2, we provide the background information and literature review, followed by the challenges faced and our experimental design in Sect. 3. In Sect. 4, we provide preliminary results and discussion, followed by a brief on future work and impact in Sect. 5.

2 Background

2.1 Defining Reflection

The term reflection when referring to a mental process is commonly understood as the action or process of thinking carefully or deeply about a particular subject, typically involving influence from one's past life and experiences.[1] More specifically in field of learning and philosophy, it was John Dewey who developed the concept of reflection as a key component of experiential learning. Dewey defined reflection as

> Active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and further conclusions to which it leads… it includes a conscious and voluntary effort to establish belief upon a firm basis of evidence and rationality (Dewey 1933, p. 9)

A key component of reflection in Dewey's writings is that reflective activity involves the perception of relationships and the connections between the parts of an experience and relating it to other experiences or beliefs from one's past life. Kolb elaborated on Dewey's work to develop a model of experiential learning cycle in which a person learns by going from concrete experience to reflective observation to abstract conceptualization and then to active experimentation to test the new concepts again into a concrete experience. According to Kolb, reflection refers to the act of associating an incoming idea with one already in the mind of the observer (Kolb and Fry 1975).

Summarizing past literature on reflection in learning, Boud et al. (2013) have defined reflection as a generic term for those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations.

Based on these definitions and writings, we can identify the phenomenon of reflection as having the following necessary and sufficient characteristics.

1. The first step towards reflection is *awareness* of a past experience.
2. Awareness of the experience is followed by *evaluation* of the experience in relation to other experiences and beliefs from a person's past.
3. The activity of reflection leads to a *new learning outcome*—an understanding or appreciation which wasn't accessible before reflection.

[1] Oxford English Dictionary, http://www.oed.com.stanford.idm.oclc.org/.

2.2 The Role of Reflection in Design Thinking

Schön (1983) took Dewey's concept of experience as interaction including the role of reflection in experiential learning and applied it to the discipline of design. Before Schön, design researchers mainly from the fields of architecture and engineering were engaged in developing prescriptive models of design activity that could guide practitioners (Bayazit 2004). Schön argued that models of technical rationality do not capture the full extent of the practice of designers. He put forward a model of reflective conversation that uses reflection-in-action as a key phenomenon to describe designers' artistry of dealing with the real-world issues that are beyond technical models. According to this model, designers engage in a conversation with the situation they are designing using activities such as sketching, prototyping etc. They develop on-the-spot hypotheses, modify the materials to represent them, and then reflect on how the modifications fit in with the situation. This could lead to new learning that helps in re-framing the situation which involves changing the perceptual meaning the situation holds for them. See Schön (1992) for a detailed description. Schön's frame of describing design struck a chord with design practitioners and researchers who had an experience of doing design work. Building on Schön's description, a number of researchers conducted studies and wrote articles describing *design reflection*.

Valkenberg and Dorst (1998) extended the study of reflection to design team interactions. Dorst and Reymen (2004) put forward seven different levels of design expertise—beginner, advanced beginner, competent, proficient, expert, master and visionary—and posit that reflection is critical to the development of higher levels of design expertise. Others such as Roozenburg and Dorst (1998) appreciated that Schön's model of reflective conversation went beyond the simplistic view of professionals applying scientific knowledge to real world problems, but at the same time criticized it for being weak and fuzzy in its definition of reflection.

Not surprisingly, design educators have incorporated reflection in their research and have developed activities and conducted studies to examine the effectiveness of reflection in learning design process (Turns et al. 1997) and design teamwork (Hirsch and McKenna 2008).

We summarize our literature review on reflection in design as follows.

1. The concept of reflection is a popular concept both in the research on understanding design activity, as well as in design education.
2. However, reflection is not well-defined as a phenomenon in design discipline.
3. The studies on reflection in design either use post-activity writings, sketches or speech recordings to collect and evaluate reflection phenomenon, or they use video for characterizing reflection-in-action. In both cases, the boundary between reflection and other forms of cognitive reasoning such as goal-oriented thinking, critical evaluation or judgment is not drawn sharply.
4. In spite of the lack of clear definition, reflection in design learning is considered a key component of developing design expertise.

2.3 Defining Rumination

In Psychology, rumination has been defined as

> a class of conscious thoughts that revolve around a common instrumental theme that recur in the absence of immediate environmental demands requiring the thought (Martin and Tesser 1996).

Thus, rumination is characterized by persistence of a thought even after the immediate stimuli is removed. When the persistent thought pertains to one's self such as feelings or memories, then the rumination is called self-reflective rumination (Nolen-Hoeksema et al. 1993). While rumination is not necessarily defined by the persistence of negative thought, prior studies have implication ruminative thinking style with increased vulnerability to depression and negative mental affect (Mor and Winquist 2002; Treynor et al. 2003).

2.4 The Role of Rumination in Design Thinking

Rumination has not directly been studied in design research. However, rumination has been shown to negatively affect creativity and problem solving. Verhaeghen et al. (2005) studied a sample of 99 undergraduate college students, using path analysis and found that self-reported past depressive symptomatology was linked to increased self-reflective rumination and rumination, in turn, was related to current symptomatology and to self-rated creative interests and objectively measured creative fluency, originality, and elaboration. The authors proposed that without a direct link between currently depressed mood and either creative interest or creative behavior, it was rumination that mediated the association between depression and creativity.

Creative problem solving is an essential characteristic of design activity. It is reasonable to hypothesize that ruminative thinking could be linked to design performance. However, the relationship between rumination and reflection and the roles that they play with respect to each other and with respect to design performance are not yet known. The experimental design we describe in the next section aims at shedding light on this relationship.

2.5 Relationship Between Reflection and Rumination

The literature review for a hypothesized or known relationship between reflection and rumination resulted in us finding the following framework by Christoff et al. (2016). The authors proposed a dynamical framework for how the mental states change over time depending on the cognitive control of deliberate constraints and non-cognitive control or automatic constraints. Christoff et al. suggested a relationship between different spontaneous mental states as shown in Fig. 1 [adapted from Christoff et al.

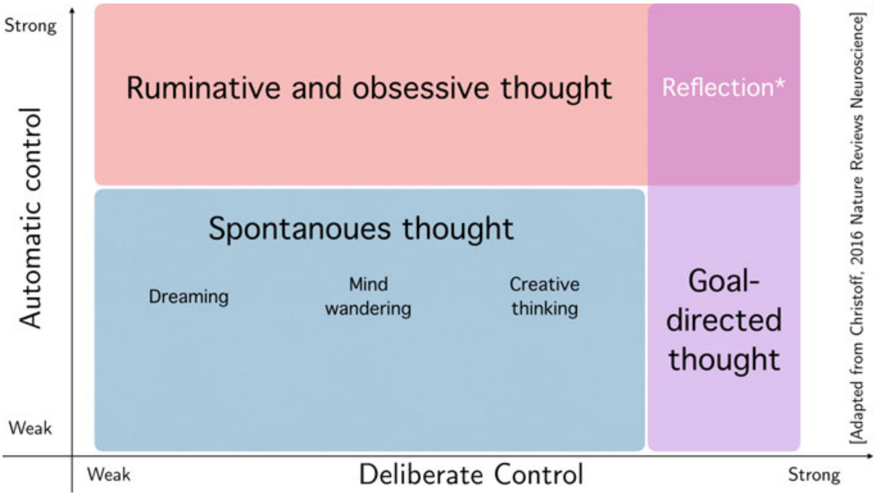


Fig. 1 Spontaneous mental states and their relationship to deliberate control and automatic constraints

(2016)].

In Fig. 1, rumination is shown as a mental state with strong automatic constraint, which means that once the state occurs it captures the attention and retains it on the ruminative thought strongly, and with weak deliberate constraint, which means that it is difficult to deliberately control ruminative thought. Our understanding of reflection is that it would fit in the upper right-hand corner where the ruminative thought overlaps with the goal directed thought. Thus, it is a mental state that has strong automatic constraint, as well as strong deliberate constraint. So, one could presume that a participant could control reflective thought much more readily than ruminative thought.

3 Our Approach

Based on the framework by Christoff et al., which highlights cognitive processes other than reflection and rumination, like goal-directed thought, creative thinking, mind wandering and dreaming, could play an important role in the study of design thinking (Fig. 2). Hence, we adapted the framework for designing current experimental study (see Fig. 2) that included mind wandering, creative thinking, goal-directed thinking and reflection/ruminative thinking. We left out dreaming since it is difficult to induce willful dreaming within the planned fMRI setup.

We developed the following experimental design to study reflection and associated thinking styles in designers participating in a team design activity. Here, we present our approach in the following three phases:

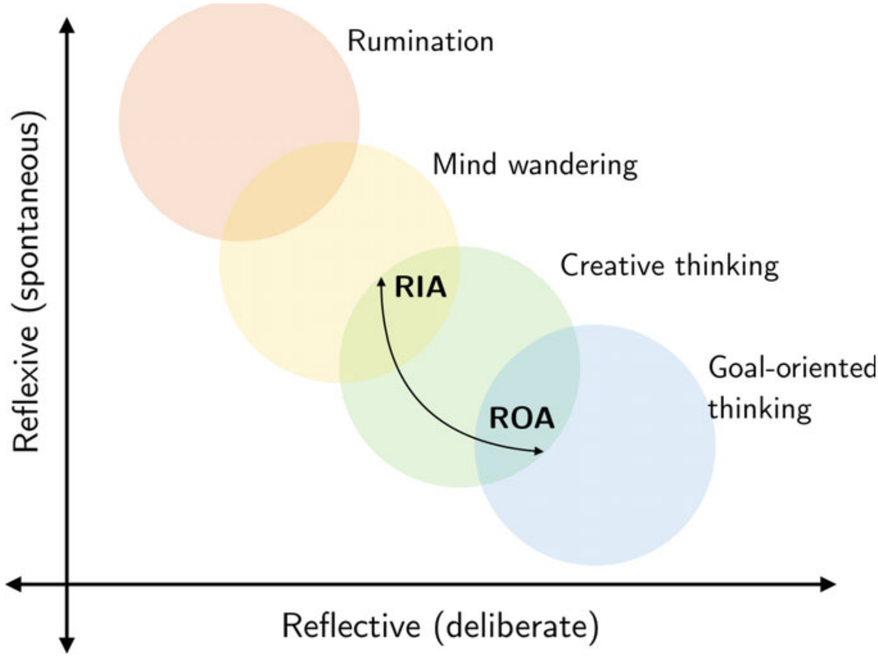


Fig. 2 Expanded conceptual framework for experiment design

3.1 The Design Team Activity Phase

In this phase, we invited participants in a triad to participate together as a team in a series of three rapid design activities. Each activity is divided in three stages—team brainstorming stage in which individuals interact and generate solution concepts together; individual prototyping stage in which individuals work on their own to prototype concepts using a box of provided materials; and team prototyping stage in which the individuals interact to synthesize their prototypes and build a common solution prototype. The following figures describes the process schematic for design team activity phase (Fig. 3).

The three design activities are video recorded in the Design Observatory setup at the Center for Design Research. Figure 4 shows the Design Observatory setup with four cameras capturing the design activity from four different perspectives.

Once the videos of design team activity are recorded, there are synced with independently recorded audio files, and are further analyzed using the Interaction Dynamics Notation (Sonalkar et al. 2013). Only five minutes of team brainstorming for each design challenge are used for IDN analysis. We conducted pilot studies (n = 6 participants) with showing the entire 15 min of video or sections of all three stages but decided to focus on brainstorming videos since they contain more interactive episodes for which IDN could be used meaningfully.

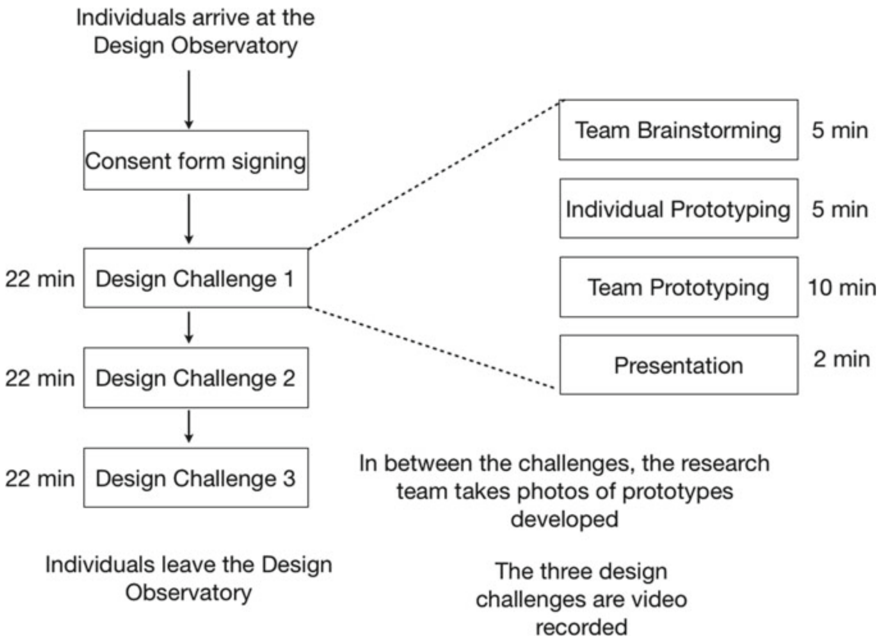


Fig. 3 Process schematic for design team activity phase

Fig. 4 Video recording setup in the Design Observatory

After conducting the IDN analysis, each brainstorming section is broken into smaller clips of various durations (10–100 s) in such a way that each clip highlights a particular aspect of team interaction—questions, humor, agreement or disagreement. These clips are shown to the subjects during the fMRI scan session.

3.2 Behavioral Assessment Phase

In the behavioral assessment phase, the participants were invited to participate individually in a series of behavioral tests. These include the following: intelligence, creativity assessments (divergent and convergent thinking, and creativity achievement), neuropsychological testing for cognitive flexibility and personality, and mind-wandering scales. Besides these tests, each participant's demographic data were also be collected in this phase.

These assessments would allow us to gain information about each participant that could be further associated with the fMRI data and/or the reflection speech data recorded as part of the scan phase.

3.3 fMRI Scan Phase

The fMRI scan phase is divided into three runs. In the first run, two video clips of his or her own team brainstorming activity or one video clip of team brainstorming activity from our pilot study were shown to each participant, during which a series of prompts showed up asking the participant to comment on the interaction type (i.e., questions, humor, agreement or disagreement). At the end of watching videos of a brainstorming activity from a design challenge, the participants were prompted to reflect on the videos they had just seen to come up with ways to improve his or her interaction with others based on it. After the scanning is over, the participants were then asked to speak out and describe what they just thought about. This speech is recorded via a mic. The speech is further analyzed using sentiment analysis for understanding the quality of reflection for each participant.

In the second run, we collected data on a number of different cognitive and affective responses to cover the expanded conceptual framework (as shown in Fig. 2). The following tasks were covered in this run—emotion, guessing, convergent thinking, theory of mind, divergent thinking, working memory, mind wandering, and visuomotor tasks.

In the third run, we collected data while participants performed the Creative Foraging task [see Hart et al. (2017)]. Here participants played a game to search for novel and valuable solutions in a large and well-defined space made of all possible shapes made of ten connected squares. Using just 10 connected squares, participants could discover categories such as digits, letters, and airplanes as well as more abstract categories. The exciting part of this game is that it allows (and measures) the amount of exploration and exploitation done by each participant and potentially tracks exploration-exploitation dynamics with the brain imaging data (Hart et al. 2018). It also allows for measuring an experimental proxy for "creative leaps" (e.g., when a new category is discovered by a participant in a non-prototypical way) (Hart et al. 2017). The Creative Foraging task has been included in the study design to be able to comment on reflection and rumination not just from the watching of videos, but also in the context of on-going exploration or exploitation phases of creative design work.

4 Challenges Faced and Discussion

The data collection for the study has been recently concluded. Here, we present some of the preliminary insights from the data. From the behavioral assessments, Figs. 5 and 6, show two teams participating in the study. The images have been modified to preserve the anonymity of the subjects.

Figures 5 and 6 show the view of all four cameras combined. For the video shown during scanning, we chose a single camera view from multiple cameras and switched one after another given by the size of the display screen inside the scanner.

Once the videos were recorded, we conducted multiple behavioral pilots for the scanner run. These consisted of prototyping the prompts that the participants would get in a scanner and then doing a test run outside the scanner in which the participants would respond to the prompts and also give feedback on what they were thinking when they saw each prompt. This helped us to refine the design of the study and develop an understanding of what might the participants be thinking when a video prompt is presented in the scanner.

> These iterative behavioral pilots for the scanner run were crucial to develop a sound experiment design that would capture the phenomenon of reflection that we aim to study, while conforming to the restrictions of the scanner environment.

Fig. 5 Team 1 Pilot performing a design challenge

Fig. 6 Team 2 Pilot performing a design challenge

These restrictions include noise, inability to move your head, small screen size and input through button-press (instead of real-time feedback). While doing these pilot runs, we realized that reflection was quite loosely defined in the design discipline and we needed to sharpen the definition as described in Sect. 2 of this chapter. The experiment design described in Sect. 3 resulted from the iterative exploration that involved behavioral pilots, literature view, and revising study frameworks.

We are currently in the process of analyzing data collected in all three phases—design team activity, behavioral assessment and fMRI scanner run—for close to 30 participants.

5 Future Work and Impact

The study we are currently implementing is oriented towards understanding the dynamics of mental states as designers reflect on their design activity. How does one enter into a reflective state? What neural patterns distinguish it from ruminative state, creative thinking, mind wandering, or goal directed thought? Is there any personality, creative, reflection, or ruminative tendency that could be identified on a behavioral assessment that correlate with the occurrence of different mental states? How are the mental state dynamics related to the actual team interaction quality or the outcome of the design task?

The investigation of these questions will help us develop a behavioral and neural model of design reflection. This when coupled with an understanding of exploration vs exploitation behavior of designers could be to develop a closed-loop behavioral and/or neural feedback system that allows designers to practice reflective thinking that could actually improve design team performance.

The first step towards creating a closed-loop reflection feedback system is to analyze the data collected in this study and develop reliable models for reflection brain dynamics. This analysis will include building brain dynamic models using a Quantified Brain Dynamics approach previously developed by our group (see Saggar et al. 2018 for details).

These brain models will be correlated with behavioral assessment data, team interaction data as measured by the Interaction Dynamics Notation, design outcomes generated in the brainstorming session recorded, the quality of reflection as noted through sentiment analysis of recorded reflection speech and exploration-exploitation dynamics during the Creativity Foraging task.

Altogether, we believe our approach will further understanding of the *interplay* between brain dynamics, design reflection, and creativity.

References

Bayazit, N. (2004). Investigating design: A review of forty years of design research. *Design Issues, 20*(1), 16–29.

Boud, D., Keogh, R., & Walker, D. (Eds.). (2013). *Reflection: Turning experience into learning*. Routledge.

Christoff, K., Irving, Z. C., Fox, K. C., Spreng, R. N., & Andrews-Hanna, J. R. (2016). Mind-wandering as spontaneous thought: A dynamic framework. *Nature Reviews Neuroscience, 17*(11), 718.

Dewey, J. (1933) *How we think*. Prometheus Books. Original work published 1910. Buffalo, N.Y.

Dorst, K., & Reymen, I. M. M. J. (2004). Levels of expertise in design education. In *DS 33: Proceedings of E&PDE 2004, the 7th International Conference on Engineering and Product Design Education*, Delft, The Netherlands.

Hart, Y., Mayo, A. E., Mayo, R., Rozenkrantz, L., Tendler, A., & Alon, U., et al. (2017). Creative foraging: An experimental paradigm for studying exploration and discovery. *PLoS ONE, 12*(8), e0182133.

Hart, Y., et al. (2018). Creative exploration as a scale-invariant search on a meaning landscape. *Nature Communications, 9,* 5411.

Hirsch, P. L., & McKenna, A. F. (2008). Using reflection to promote teamwork understanding in engineering design education. *International Journal of Engineering Education, 24*(2), 377.

Kolb, D. F., & Fry, R. R.(1975).Towards an applied theory of experiential learning. *Theories of group processes*, 33–58.

Martin, L. L., & Tesser, A. (1996). Some ruminative thoughts. In R. Wyer Jr. (Ed.), *Ruminative thoughts* (Vol. 11, pp. 1–47)., Advances in social cognition Hillsdale, NJ: Erlbaum.

Mor, N., & Winquist, J. (2002). Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin, 128,* 638–662.

Nolen-Hoeksema, S., Morrow, J., & Fredrickson, B. L. (1993). Response styles and the duration of episodes of depressed mood. *Journal of Abnormal Psychology, 102,* 20–28.

Roozenburg, N. F., & Dorst, K. (1998). Describing design as a reflective practice: Observations on Schön's theory of practice. In *Designers* (pp. 29–41). London: Springer.

Saggar, M., et al. (2018). Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature Communications, 9,* 1399.

Schon, D. A. (1983). *The reflective practicioner: How professionals think in action* (Vol. 1). New York: Basic books.

Schön, D. A. (1992). Designing as reflective conversation with the materials of a design situation. *Knowledge-Based Systems, 5*(1), 3–14.

Sonalkar, N., Mabogunje, A., & Leifer, L. (2013). Developing a visual representation to characterize moment-to-moment concept generation in design teams. *International Journal of Design Creativity and Innovation, 1*(2), 93–108.

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research, 27,* 247–259.

Turns, J., Newstetter, W., Allen, J. K., & Mistree, F. (1997). Learning essays and the reflective learner: Supporting reflection in engineering design education. In *Proceedings of ASEE Annual Conference and Exposition*, Milwaukee, WI.

Valkenburg, R., & Dorst, K. (1998). The reflective practice of design teams. *Design Studies, 19*(3), 249–271.

Verhaeghen, P., Joorman, J., & Khan, R. (2005). Why we sing the blues: the relation between self-reflective rumination, mood, and creativity. *Emotion, 5*(2), 226.

Whiteman, R. C., & Mangels, J. A. (2016). Rumination and rebound from failure as a function of gender and time on task. *Brain Sciences, 6*(1), 7.

Tools to Support Design Thinking Practices

Prototyper: A Virtual Remote Prototyping Space

Matthias Wenzel and Christoph Meinel

Abstract Collaborative virtual environment groupware is—despite of notable research efforts over several years—still not common in users' workplaces. Reasons are high costs of engaging in collaboration next to the loss of information and capabilities that people enjoy in co-located settings. Low-fidelity prototyping is a way for co-located teams to create joint understandings and to gather feedback in early design stages. When it comes to geographically dispersed teams, dedicated tools are required that help to fulfill tasks at hand, while enabling team members as much as possible to apply working modes known from co-located settings. We present a web browser-based collaborative virtual environment that supports the joint real time creation of three-dimensional low-fidelity prototypes. It is a cross-platform application that runs on a multitude of hardware devices. While focusing on usage with virtual reality hardware, users may also freely participate when there are only traditional input and output devices available. The system provides enhanced awareness through visual remote user embodiment combined with spatial audio communication.

1 Introduction

Remote collaboration tools are well-established in present-day work environments. However, collaborative virtual environment groupware is still not common at workplaces. One reason is the high costs of engaging in collaboration: Systems require a complex setup and usually focus on a small, homogeneous spectrum of hardware and software, excluding participants who do not satisfy all technical requirements. The loss of information and capabilities people are used to when working in co-located settings is another reason for the low utilization rate of collaborative virtual environment groupware systems.

M. Wenzel (✉) · C. Meinel
Hasso Plattner Institute for Digital Engineering, Prof.-Dr.-Helmert Straße 2-3,
14482 Potsdam, Germany
e-mail: matthias.wenzel@hpi.de

C. Meinel
e-mail: meinel@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_11

Collaboration is a fundamental aspect of Design Thinking (DT). Team members generate ideas and communicate these to their teammates and other stakeholders. Ideas can be expressed in a tangible way, such as in three-dimensional hand-crafted prototypes using haptic materials (e.g. paper or bricks). As a dedicated representation of an evolving design (Houde and Hill 1997), a prototype serves as an illustration of people's ideas and to externalize implicit knowledge (Buxton 2007). DT's team-based approach is often applied in co-located settings.[1] Hence, the provided tools and materials mostly aim at supporting co-located teams. However, working at the same location is not always possible for all participants of a DT team. Dedicated software tools can help teams to work and prototype together over distances.

In this chapter, we present a web browser-based collaborative virtual environment for supporting geographically distributed design teams in their joint creation of three-dimensional low-fidelity prototypes. The application provides a shared 3D workplane surrounded by remote users' avatars. All participants can jointly create and modify 3D shapes while seeing and hearing each other in real time. While focusing on virtual reality (VR) hardware, such as head-mounted display and respective controllers, our cross-platform application can also be operated by mouse or touch on mobile as well as on traditional desktop devices.

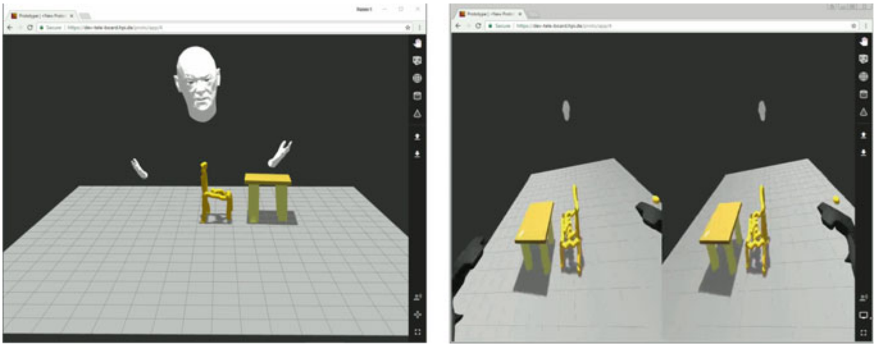
2 Prototyper

Figure 1 shows the application with two remote participants. The user at location A views the provided prototyping space on a traditional computer screen, whereas location B's user wears a head-mounted display that shows a three-dimensional stereoscopic visualization of the prototyping space. The remote participant can equally interact with generated 3D artifacts using available input devices, such as mouse (location A) or dedicated 3D controllers (location B). At the same time, both users can see each other's avatar, its current position and viewing direction.

As a valuable part for distributed collaboration (Tang 1991; Whittaker et al. 1993), a shared workspace is required for jointly manipulating 3D artifacts in our application's virtual prototyping space. Its conceptual setup is depicted in Fig. 2. The virtual prototyping space in Fig. 2b is modelled on the table setup schema shown in Fig. 2a.

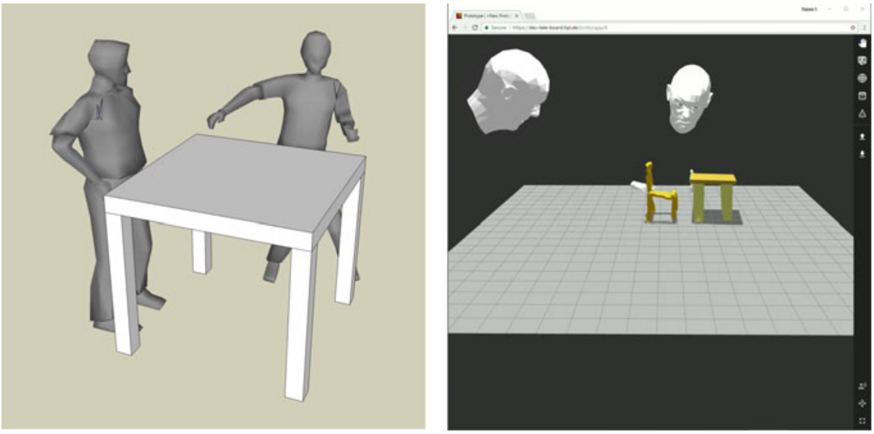
Remote users' virtual embodiments are placed around a table representing a workplane that serves as the shared workspace where 3D artifacts can be created, modified and assembled to more complex structures. The proportions regarding participants' body size and the table's dimension shown in the schema are also maintained in the setup's virtual counterpart. Users should be within arm's reach of the 3D objects on

[1] https://hpi.de/en/school-of-design-thinking/design-thinking/—Accesed Jan. 2019
https://dschool.stanford.edu/about/—Accesed Jan. 2019
When we speak about DT, we are referring to the way DT is commonly applied and taught at Hasso Plattner Institute's School of Design Thinking and d.school Stanford.



Location A - Computer screen view Location B - Virtual reality stereoscopic view

Fig. 1 The Prototyper web application being used simultaneously at two different locations. Users operate and view the application with traditional hardware, such as a computer screen and mouse (left) or with dedicated virtual reality hardware (right)



(a) Setup schema (b) Applied schema in Prototyper's implementation

Fig. 2 Prototyper table setup

the workplane. This is important for the application's interaction concept described below.

Given this setup, our application implements three distinct types of spaces proposed by Buxton 2009: a person space of the remote participants' avatars, a task space of involved 3D artifacts on the workplane and a reference space for pointing. Participants are aware of others' activities since they can see who is present, where users look, and which part of the model they are working on.

2.1 3D Modeling Using a Basic Construction Kit

With Prototyper, we aim to build volumetric low-fidelity prototypes (e.g. the prototypes shown in Fig. 3[2]). These prototypes only provide key elements of the underlying visual concept with a rather low demand regarding level of detail (Walker et al. 2002) and user skills (Babich 2017).

These aspects are reflected by our application's 3D model construction and modification concept. Prototypes consist of a small set of building blocks, i.e. basic 3D shapes that can be transformed, colorized and textured. Figure 4 shows the four basic shapes our system offers. All shapes can be composed to more complex structures using Boolean set operators-based *Constructive Solid Geometry* (CSG) (Foley et al. 1990). The technique's results are shown in Fig. 4b: as a first step, the three scaled and rotated *cylinder* shapes in the middle are combined utilizing a Boolean `union` operation; in a second step, the shape on the right is created by applying a `difference` operation on the left sided *cube* and the combined *cylinder* shapes.

The combination of basic shapes and repetitive CSG allows the creation of complex 3D models. As described by Wenzel et al. 2016, users are able to resemble the real-world prototypes shown in Fig. 3. However, during initial user tests we

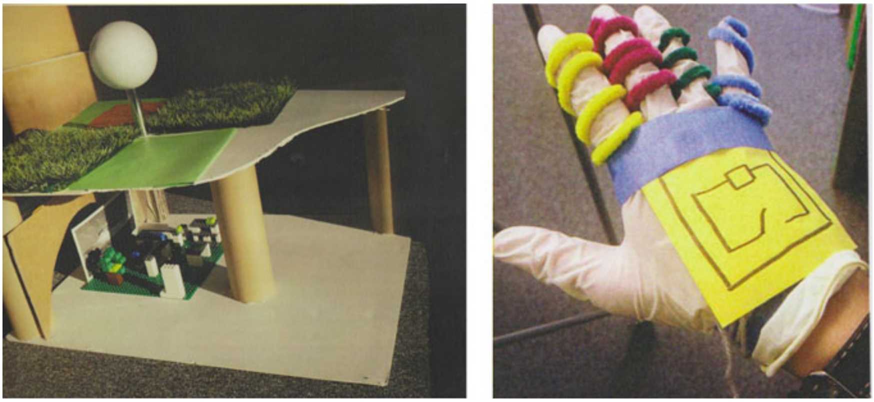
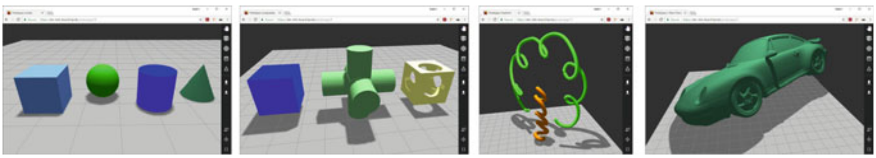


Fig. 3 Examples of low fidelity prototypes



(a) Simple (b) Composites (c) Freeform (d) External

Fig. 4 Prototyper's set of basic shapes

[2] Actual prototypes built during DT projects at Hasso Plattner Institute's School of Design Thinking.

learned that the application of CSG requires some experience from users, especially for beginners, in constructing 3D models. Hence, we added freeform shapes (see Fig. 4c) whose creation is easier and more intuitive for beginners. Conceptually, creating Prototyper's freeform shapes in 3D relates to drawing lines with a marker on a sheet of paper in 2D. Our freeform approach is similar to the hand-held physical sketching device described by Agrawal et al. 2015, which proved to be useful for creative exploration. Creating freeform shapes in Prototyper is only possible in VR mode with respective 3D controllers, providing well-defined, continuous 3D positional data. Though being generally input method agnostic, this is our application's only limitation regarding 3D content generation.

In the case that Prototyper is not initially used for creating 3D models from scratch, the system allows importing external 3D files and using the corresponding models just as any other shape within the system. An example for such an external model is shown in Fig. 4d. Even if it would not be intended to further modify an external model, Prototyper can serve as a tool for jointly viewing and discussing later stage high-fidelity prototypes.

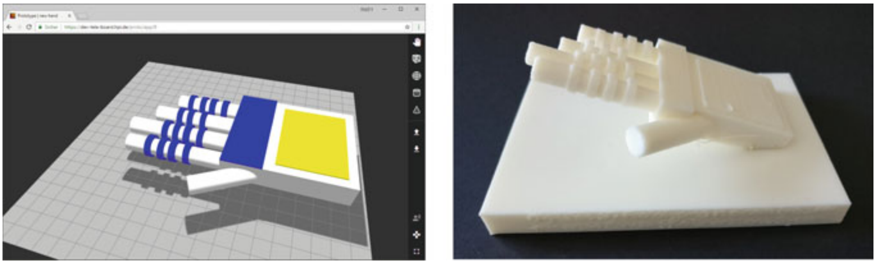
2.2 Interfaces to the Analog World—Import and Export of Physical and Virtual Prototypes

Prototyper covers a specific use case. In order to preserve generated content and to embed Prototyper in larger contexts within users' day-to-day work, a "bridge" to other already existing software systems is needed. Importing external 3D files into our system allows building upon existing 3D models (e.g. digital models created with a *Computer Aided Design* (CAD) tool). For the other direction, models created with Prototyper can in turn be exported into 3D files.

The exchange interfaces allow a connection to physical objects. Though not widespread at the moment, there are research efforts toward digitizing three-dimensional objects with a 3D scanner built into smaller devices, such as smartphones (Naegeli 2013; Stoller-Conrad 2015). In order to gather feedback on a digital 3D model, a 3D printer can be used to create a physical object (see Fig. 5b).

2.3 Awareness Through Audio-Visual Remote User Embodiment

In general communication scenarios, speech is an important instrument for explicitly exchanging information and for gathering evidence that a message has been understood as intended (Clark and Brennan 1991). However, in co-located, face-to-face settings, people's communication is also based on rather implicit sources of information, such as the views of others' faces, bodies, and actions; views of the task



(a) Virtual Prototyper model (b) 3D printed model

Fig. 5 Virtual 3D model created with Prototyper, and the model's physical, 3D printed version

objects; and views of the environment (Kraut et al. 2003). These visual cues help provide awareness of other group members, which is crucial for successful collaboration (Gutwin and Greenberg 2004; Dourish and Bellotti 1992), especially when the design and use of artifacts are involved (Poppe et al. 2013).

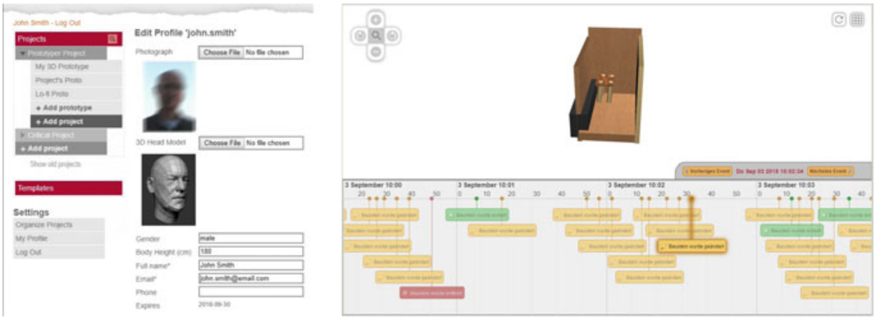
Within collaborative virtual environments, it is challenging to coordinate and manage the actions and intentions of users (Domingues et al. 2010). Three-dimensional user embodiment within virtual environments is an approach for addressing these challenges. Benford et al. 1995 consider user embodiment a key issue for collaborative virtual environments:

> …without sufficient embodiment, users only become known to one another through their (disembodied) actions; one might draw an analogy between such users and poltergeists, only visible through paranormal activity.

Gutwin and Greenberg (1999) and Benford et al. (1995) identified a list of design issues that are important for awareness in real time groupware systems and collaborative virtual environments, respectively. These issues can be split into categories for providing information about (1) who users are working with, (2) what others are doing, (3) where they are working, (4) when events happen, and (5) how those events occur (Gutwin and Greenberg 1999).

Within Prototyper, we seek to address all of these issues based on 3D user embodiment. Remote participants' avatars, as shown in Fig. 2b, provide information about the presence of a person, the location, gaze and field-of-view. The 3D artifacts on the workplane, together with the virtual hands of a remote user inform about what artifacts are modified by whom. The scale of the visual appearance of remote avatars, workplane and its objects, applies to the real-world setting shown in Fig. 2a. This way, a user can estimate a remote participant's reach within the virtual space.

A limitation within Prototyper is that the system does not provide an actual live image of a remote participant. However, in order to provide information about the identity of a remote user we use additional metadata. Prototyper is a web-based system that includes—beside the Prototyper application itself—a web portal that provides user management, content organization and access control. This means, every Prototyper user has a user account connected with a profile (see Fig. 6a) where



(a) User profile in the web portal.

(b) Prototyper history browser. Users can navigate to any point in time of a prototype's course of development using the timeline.

Fig. 6 Prototyper system web portal

he/she can upload a profile photo used within the web portal but, more importantly, can also upload a 3D file to be used as a personal avatar within Prototyper.

With the help of the person's gender and body height, we calculate the size and position of the 3D avatar (Medlej 2013). When no avatar file is specified, then a default avatar is displayed within Prototyper based on the provided gender. This way, our system provides at least avatar information that lets users distinguish between different remote participants helping to assign actions to specific persons.

Information regarding when and how events happen is currently not visualized directly within Prototyper. However, this information is stored automatically on a server storage for later use.

When using Prototyper, the people are connected via audio so they can talk to each other. In order to help users to distinguish who is actually talking, we provide a visual hint. The respective remote avatar gets a special color whenever talking. Visual cues regarding the user and remote participants as well as their activities in relation to the workspace materials are crucial but are oftentimes not sufficient for a sense of presence (Büscher et al. 2001). Especially when users are looking in another direction or the remote participant is not within a user's field-of-view it can become difficult to know the speaker's identity. In co-located settings people can locate another person by the sound of his/her voice. Within our system, we try to resemble this kind of out-of-sight localization. Based on the positions of the local user and the remote participant in the virtual space—this data is exchanged among all locations in real time—the remote user's voice is adapted to "hear" his/her position.

2.4 A Web Browser-Based Cross-Platform Application for Immediate Access

The costs for a user to engage in a collaboration are a critical factor for the success of a remote collaboration system, i.e. that the users' effort be kept to a minimum (Gutwin et al. 2008; Kraut et al. 2002).

User interface and user experience issues have to be considered when designing a remote collaboration system. However, in order to get to the point of experiencing a system, users have to actually use it, which might become problematic when it means telling the other person: "OK, remember what you wanted to say while I go and find a room, power up the system and install the software."

Prototyper is a web browser-based application. This has two major advantages: (1) web browsers are available and mostly pre-installed on almost all user devices from mobile to traditional desktop hardware systems and (2) people know web browsers from their daily life; they know how to operate them and the corresponding paradigms.

Hence, Prototyper's only requirement is a web browser. An installation of additional software is not necessary. This applies to operating Prototyper with mouse or touch input. For the intended usage with virtual reality hardware, additional (mostly driver-) software installation is required. In the current version of our system, we support *HTC Vive*[3] as virtual reality hardware. Prototyper is designed in such a way that other VR hardware can be added easily, e.g. mixed reality systems and mobile phones.

From a user perspective, the Prototyper system consists of the 3D modeling application and a web portal that serves as an entry point and administration interface for the system. Users can manage projects and associated prototypes in order to organize their work and control access rights. The menu on the left in Fig. 6a shows three prototypes contained in a project. User accounts are assigned to projects and therefore gain access to the projects' prototypes.

The Prototyper application can be started from the web portal by any user that is assigned to the prototype's project. All participants get connected automatically when the Prototyper application is started. From that point, all user interactions are synchronized among all participants in real time via our central collaboration server. The server does not only relay the synchronization messages but also stores the content data. When working with Prototyper, the data is stored automatically so users do not have to press a "save" button. This way, the latest state is shown when starting Prototyper. However, since all of a prototype's content data is stored from the beginning, this data can be viewed via a dedicated history browser (see Fig. 6b) allowing users to navigate through a prototype's course of development, creating a branch from any point in time if necessary.

[3]https://www.vive.com/de/product/—Accessed Jan. 2019.

2.5 Interaction

The ability to manipulate objects within virtual environments, such as virtual reality, is a defining feature for such systems (Bowman et al. 1997). This rises the need for interface and interaction techniques focusing on spatial input in a physical three-dimensional context (Bowman et al. 2012).

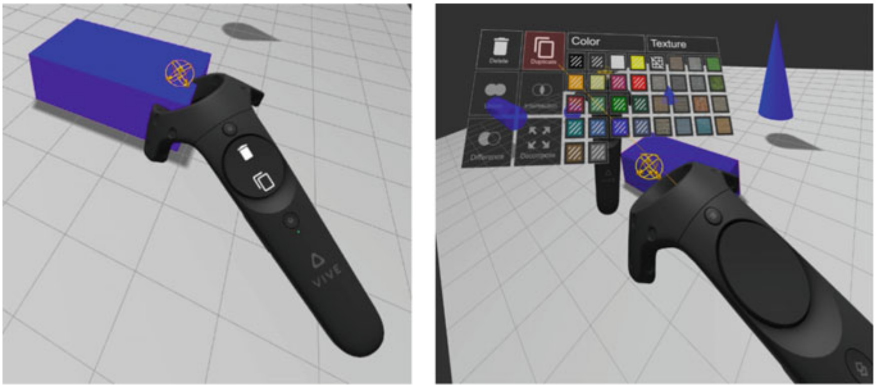
There are three basic user interaction tasks within virtual environments: *navigation* or *travel*, *selection* and *manipulation* (Bowman et al. 1997; Bowman and Hodges 1999; Bowman et al. 2001). There are different technical approaches for each of these tasks. Each of them can be distinguished with regard to the "objective degree of exactness with which real-world interactions can be reproduced in an interactive system" (McMahan 2011). High levels of this naturalism can enhance performance and the overall user experience (Bowman et al. 2012).

2.5.1 Selection and Manipulation

Raycast and arm extension techniques are common approaches for the object selection and manipulation tasks (Bowman and Hodges 1997). These techniques usually originate from the need to reach and interact with objects that are further away and not within arm's reach. Without such techniques, users would need to change their avatar's position within the virtual space, i.e. they would have to travel. Compared to other techniques, raycasting and arm extension methods' level of naturalism is lower since there is no exact, direct mapping of user's arm and the resulting movement in the virtual world. A common, and more natural technique, is called *simple virtual hand* (Bowman et al. 2004). Here, users control a virtual hand directly mapped to user's real hand movements. For selecting and manipulating, users touch a virtual object, in a way that parallels how users interact with real-world objects. This technique requires the user to be within arm's reach. The intended interaction space within Prototyper is basically the workplane, whose size is designed to be within arm reach. Thus, we utilize the simple virtual hand technique in our system. Users' virtual hands are represented by real-world sized virtual controllers that the user has in his/her hands. Figure 7a shows the controller touching a virtual object.

When there are many objects in a smaller area it becomes difficult to distinguish which object is actually being touched. Therefore, Prototyper provides a visual cue by means of a slight color change in the touched object. Interacting with virtual objects is also challenging for users due to the lack of haptic feedback (Mine et al. 1997). In order to provide, at least a small part of such feedback, a force feedback motor causes a slight controller vibration whenever an object is touched. Users preferred our visual/haptic hand approach over a simpler raycast version that we had implemented first.

Moving and rotating virtual objects with the simple hand approach is quite straightforward: when an object is grabbed, it is attached to the user's hand, or controller respectively, so that all translation and rotation transformations by the



(a) Touching and grabbing a virtual object.

(b) Changing object's state by setting attributes using a context menu.

Fig. 7 Interaction techniques: **a** Simple virtual hand and **b** Raycasting. When touching a virtual object, a shortcut menu is shown on the virtual controller for fast reachability of common options (**a**). Raycasting is used to select options on a controller aligned context menu (**b**)

controller are applied on the object in the same way. A different approach is necessary for changing the size of a virtual object. Our solution for scaling is to divide horizontal, vertical and uniform scaling based on the controller orientation and the number of involved controllers. Holding a special scale button on one controller causes the grabbed object to scale in horizontal dimensions by controller's amount of translation when the controller is oriented horizontally. Vertical scaling is realized with a vertical controller orientation. For a uniform scale, the user just grabs an object with both hands/controllers and pulls or squeeze it to increase or decrease its size.

2.5.2 System Control Using Menus

With Prototyper, objects can be created, transformed or changed in their visual appearance. This functionality is provided by different tools and options within the system. The task of changing a virtual system's state or mode of interaction is called *system control* (Bowman and Wingrave 2001). Changing object's visual attributes or switching between different tools is realized with the help of different menus. Figure 7b shows a menu for changing an object's visuals. The menu is displayed on the user's hand. It is also attached to the controller so that it moves and rotates with it. Since the selection icons are smaller, touching these is difficult. Furthermore, selecting an option is just a "click". We therefore provide a ray originating from the other hand to make this binary click selection.

Prototyper provides three different interaction modes: (1) object transformation (2) object creation and (3) freeform drawing. Switching between these modes is shown in the left picture of Fig. 8. Pressing the mode selection button on a controller

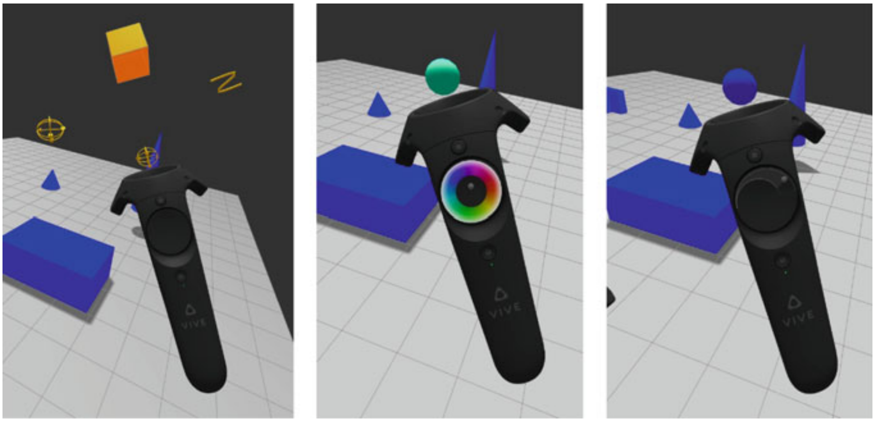


Fig. 8 Menus for switching between different modes of interaction and setting options and parameters

shows three working mode symbols that users can select by rotating the controller around its longitudinal axis pointing to the desired tool symbol.

The touch sensitive area of the controllers are used to dynamically display and select different options, such as color and thickness of freeform drawings (see middle and right picture of Fig. 8) or shortcut options for object duplication and deletion when touching an object (see Fig. 7a).

2.5.3 Travel

As already mentioned, the interaction area within Prototyper is relatively small. Travelling should not be necessary since objects are usually within arm's reach of the user. However, since research shows that physical turning and walking can enhance spatial orientation and movement understanding (Bowman et al. 2012), users are supposed to physically walk in order to change their avatar's position when using Prototyper. This requires a precise *Six degrees of freedom* (6DoF) tracking within a larger area. Given modern VR systems, both requirements are fulfilled providing precise tracking space of e.g. 5-meter diagonal.[4]

2.5.4 Mouse and Touch Interaction

An important factor is the support of heterogeneous hardware in regard to user input devices. This way, users who do not have dedicated VR hardware can take part in Prototyper's collaborative virtual environment.

[4] https://www.vive.com/us/support/vive/category_howto/what-is-the-recommended-space-for-play-area.html, Accessed Apr. 2018.

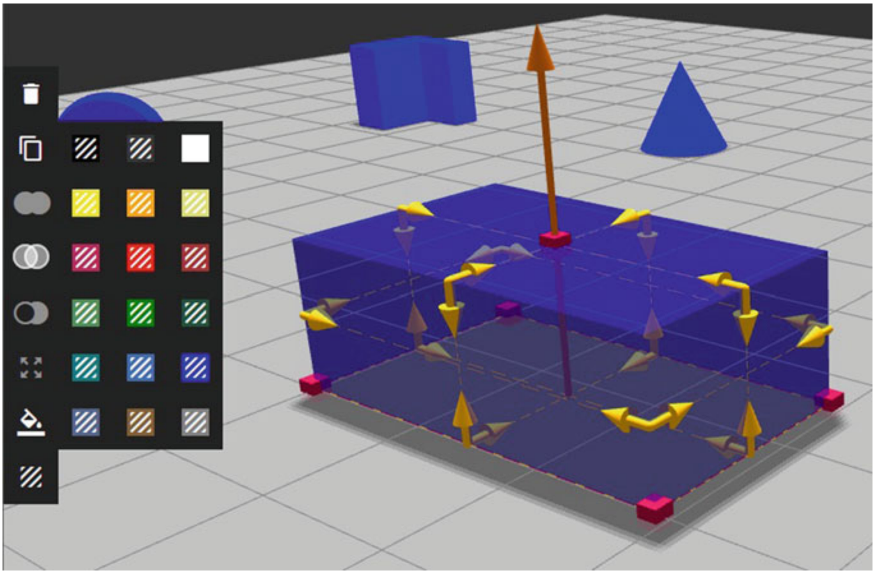


Fig. 9 User interface for mouse- and touch-based object transform and styling

Hence, Prototyper provides a "fallback" mechanism for the aforementioned user interaction tasks. Figure 9 shows the user interface for object transforms and visual attributes change.

This approach enables synchronous collaboration with different hardware setups.

3 Conclusion and Future Work

In this chapter, we presented Prototyper, a web browser-based collaborative virtual environment that supports the joint real time creation of three-dimensional low-fidelity prototypes. It is a cross-platform application that runs on a multitude of hardware devices. While focusing on a usage with virtual reality hardware, users are free to participate when there are only traditional input and output devices available. The system provides enhanced awareness through visual remote user embodiment combined with spatial audio communication.

In future work, we want to focus on mobile devices, such as smartphones, and the user interactions in VR when there are either no or only basic input devices available.

References

Agrawal, H., Umapathi, U., Kovacs, R., Frohnhofen, J., Chen, H. T., Mueller, S., & Baudisch, P. (2015). Protopiper: Physically sketching room-sized objects at actual scale. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology—UIST '15* (pp. 427–436) (2015). https://doi.org/10.1145/2807442.2807505.

Babich, N. (2017). *Prototyping 101: The difference between Low-Fidelity and high-fidelity prototypes and when to use each. Adobe blog*. Accessed April 2018. (November 2017). https://theblog.adobe.com/prototyping-difference-low-fidelity-high-fidelity-prototypes-use/.

Benford, S., Bowers, J., Fahlén, L. E., Greenhalgh, C., & Snowdon, D. (1995). User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)* (pp. 242–249). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co. https://doi.org/10.1145/223904.223935.

Bowman, D. A, & Hodges, L. F. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics (I3D '97)* (pp 35–38). New York, NY, USA: ACM. https://doi.org/10.1145/253284.253301.

Bowman, D. A, & Hodges, L. F. (1999). Formalizing the design, evaluation, and application of interaction techniques for immersive virtual environments. *Journal of Visual Languages & Computing*, *10*(1), 37–53.

Bowman, D. A., Koller, D., & Hodges, L. F. (1997). Travel in immersive virtual environments: An evaluation of viewpoint motion control techniques. In *Virtual Reality Annual International Symposium* (pp. 45–52), IEEE.

Bowman, D. A., Kruijff, E., LaViola, J. J., Jr., & Poupyrev, I. P. (2004). *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional.

Bowman, D. A., McMahan, R. P., & Ragan, E. D. (2012). Questioning naturalism in 3D user interfaces. *Communications of the ACM, 55*(9), 78–88. https://doi.org/10.1145/2330667.2330687.

Bowman, D. A, & Wingrave, C. A. (2001). Design and evaluation of menu systems for immersive virtual environments. In *Proceedings IEEE Virtual Reality 2001* (pp. 149–156). IEEE. https://doi.org/10.1109/VR.2001.913781.

Büscher, M., O'Brien, J., Rodden, T., & Trevor, J. (2001). "He's behind you": The experience of presence in shared virtual environments. In *Collaborative Virtual Environments: Digital Places and Spaces for Interaction* (pp. 77–98). London: Springer. https://doi.org/10.1007/978-1-4471-0685-2_5.

Buxton, B. (2007). *Sketching user experiences: Getting the design right and the right design*. Morgan Kaufmann.

Buxton, B. 2009. Mediaspace—Meaningspace—Meetingspace. In S. Harrison (Ed.), *Media space: 20+ years of mediated life, number 2009 in computer supported cooperative work* (pp. 217–231). London: Springer, https://doi.org/10.1007/978-1-84882-483-6_13.

Clark, H. H., Brennan, S. E., et al. (1991). Grounding in communication. *Perspectives on socially shared cognition* 13 (pp. 127–149).

Domingues, C., Davesne, F., Mallem, M., & Otmane, S. (2010). Collaborative 3D interaction in virtual environments: A workflow-based approach. In *Virtual reality*. InTech. https://doi.org/10.5772/13013.

Dourish, P., & Bellotti, V. (1992). Awareness and coordination in shared workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work (CSCW '92)* (pp. 107–114). New York, NY, USA: ACM. https://doi.org/10.1145/143457.143468.

Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1990). *Computer graphics: Principles and practice* (2nd ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co. Inc.

Gutwin, C., Greenberg, S. (1999). The effects of workspace awareness support on the usability of real-time distributed groupware. ACM Transactions on Computer-Human Interaction *6*(3), 243–281 (1999). https://doi.org/10.1145/329693.329696.

Gutwin, C., & Greenberg, S. (2004). The importance of awareness for team cognition in distributed collaboration. In *Team cognition: Understanding the factors that drive process and performance* (pp. 177–201). Washington: American Psychological Association. https://doi.org/10.1037/10690-009.

Gutwin, C., Greenberg, S., Blum, R., Dyck, J., Tee, K., & Mcewan, G. (2008). Supporting informal collaboration in shared-workspace groupware. *Journal of Universal Computer Science*, *14*, 1411–1434 (2008). https://doi.org/10.3217/jucs-014-09-1411.

Houde, S., & Hill, C. (1997). What do Prototypes Prototype? In *Handbook of Human-Computer Interaction. Vol. 2* (pp. 367–381).

Kraut, R. E., Fussell, S. R., & Siegel J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human–Computer Interaction 18*(1), 13–49. https://doi.org/10.1207/S15327051HCI1812_2.

Kraut, R. E., Fussell, S. R., Brennan, S. E., & Siegel, J. (2002). Understanding effects of proximity on collaboration: Implications for technologies to support remote collaborative work. *Distributed work,* 137–162.

McMahan, R. P. (2011). Exploring the effects of higher-fidelity display and interaction for virtual reality games. Ph.D. Dissertation, Virginia Tech.

Medlej, M. (2013*). Human Anatomy Fundamentals: Basic Body Proportions*, Accessed April 2018. (November 2013). https://design.tutsplus.com/articles/human-anatomy-fundamentals-basic-body-proportions–vector-18254.

Mine, M. R., Brooks, F.P., Jr., & Sequin, C. H. (1997). Moving objects in space: Exploiting proprioception in virtual-environment interaction. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)* (pp. 19–26). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co. https://doi.org/10.1145/258734.258747.

Naegeli, C. (2013). Your Smartphone as a 3D Scanner. Eidgenössische Technische Hochschule (ETH) Zurich. Accessed April 2018. (December 2013). https://www.ethz.ch/en/news-and-events/eth-news/news/2013/12/your-smartphone-as-a-3d-scanner.html.

Poppe, E., Brown, R., Recker, J., & Johnson, D. (2013). Improving remote collaborative process modelling using embodiment in 3D virtual environments. In *Proceedings of the Ninth Asia-Pacific Conference on Conceptual Modelling* (vol. 143, pp 51–60). (2013) http://dl.acm.org/citation.cfm?id=2527203.

Stoller-Conrad, J. (2015). New camera chip provides superfine 3-D resolution. In *California Institute of Technology (Caltech)*. Accessed April 2018. (April 2015). https://www.caltech.edu/news/new-camera-chip-provides-superfine-3-d-resolution-46425.

Tang, J. C., (1991). Findings from observational studies of collaborative work. International Journal of Man-Machine Studies, *34*(2), 143–160. https://doi.org/10.1016/0020-7373(91)90039-A.

Walker, M., Takayama, L., Landay, J. A. (2002). High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (vol. 46(5), pp. 661–665) (2002). https://doi.org/10.1177/154193120204600513.

Wenzel, M., Klinger, A., & Meinel, C. (2016). Tele-Board Prototyper—distributed 3D modeling in a web-based real-time collaboration system. In *2016 International Conference on Collaboration Technologies and Systems (CTS)* (pp. 446–453). Orlando, FL. https://doi.org/10.1109/cts.2016.0084.

Whittaker, S., Geelhoed, E., & Robinson, E. (1993). Shared workspaces: how do they work and when are they useful? *International Journal of Man-Machine Studies*, *39*(5), 813–842. https://doi.org/10.1006/imms.1993.1085.

Investigating Active Tangibles and Augmented Reality for Creativity Support in Remote Collaboration

Mathieu Le Goc, Allen Zhao, Ye Wang, Griffin Dietz, Rob Semmens and Sean Follmer

Abstract Physical manipulation is a key part of externalizing representations of knowledge and the creative process. However, contemporary tools for remote collaboration ignore physical manipulation and the haptic modality. We are interested in exploring remote physical manipulation in the context of ideation and brainstorming. Augmented Reality provides much of the benefits of spatial representation of remote participants, yet AR does not allow for rich physical manipulation and haptic feedback. Thus, we propose to use pairs of multi-robot system to provide synchronized haptic proxies in conjunction with the AR system. These small, tangible robots can be used directly as handles for digital models. We share insights gathered during experimentation to help design platforms combining AR and actuated tangibles, and present several application scenarios to illustrate their potential for remote collaboration.

1 Introduction

With the recent democratization of online platforms and services to facilitate collaboration at a distance, it is becoming increasingly easier to work with people across the planet, sometimes without ever meeting them. From Google GSuite to Microsoft Office 365, Skype, Apple FaceTime, and Cisco Webex, long is the list of platforms and tools enabling and empowering users to work and collaborate remotely. We even see design software such as Solidworks, Rhinoceros, and 3DS Max now supporting live collaboration, allowing engineers and designers to concurrently edit 3D models. Yet, while this is suitable for a large number of tasks, many contexts still require more direct styles of collaboration. In highly collaborative situations such as brainstormings or prototype evaluations, physical co-location encourages designers to share, communicate and discuss.

M. Le Goc (✉) · A. Zhao · Y. Wang · G. Dietz · R. Semmens · S. Follmer
SHAPE Lab, Stanford University, 524 Duena St., Stanford, CA 94305, USA
e-mail: mlegoc@stanford.edu

R. Semmens
Naval Postgraduate School, Monterey, CA, USA

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_12

Existing collaborative tools often fall short when it comes to communication channels. They are remain limited in resolution and bandwidth, failing to capture the rich and diverse subtleties of our communication capabilities. For instance, the diversity of gestures and expressions allows us to convey complex and intricate details that could be challenging to share with words, such as referring to a specific parts a features of a design or exploring textures of a material. Design thinking practices leverage more and more physical artifacts to offload cognition and embody ideas with physical objects.

Successful collaboration relies on both verbal and non-verbal communication skills. Designers especially make use of non-verbal communication, using gesture, sketching and physical artifacts to ground conversation. In shared collaboration Buxton (Buxton 2009) outlined how space can be partitioned into three distinct areas: person space, task space, and reference space. Traditional video conferencing tools do not support all of these modalities, while research in shared media spaces has sought to support these different spaces in the digital world. However, providing a physical representation of task space shared across distances has remained a challenge. Our belief is that designers require a collaborative tool that supports all three of Buxton's spaces while allowing for tangible collaboration across distances.

In this work, we seek to support physical remote collaboration in the context of design thinking and creative spatial tasks by leveraging the use of augmented reality (AR) for rendering of remote participants and a pair of synchronized actuated tangible interfaces to share users' physical workspace or task space. We build on the open-source Zooids (Le Goc 2016) platform, a tabletop swarm user interface which is comprised of many actuated tangible pucks. Our system supports person space, task space, and reference space (Buxton 2009) towards the goal of increasing perceived co-presence and enabling rich nonverbal cues, while leveraging the benefits of physical manipulation in design and spatial tasks. We are specifically interested in the following research questions:

- Can physicality better support remote collaboration?
- How can active tangibles improve remote collaborative manipulation tasks?

2 Background

2.1 Computer Supported Cooperative Work

Researchers have investigated computer supported collaborative work (CSCW) in various contexts, such as sharing physical documents using video feeds on a screen, as in TeamWorkStation (Ishii and Miyake 1991), or projected, as in Video Draw [30]. Video Whiteboard explored wall-scale shared workspaces, while providing feedback on remote user's presence (Tang and Minneman 1991). Clearboard (Ishii and Kobayashi 1992) went further by allowing for proper gaze estimation of remote users. More recently, these techniques have been applied to applications, such as collaborative website development (Everitt et al. 2003), remote board games (Wilson and Robbins 2007) and family communications (Yarosh et al. 2013).

2.2 TUI for Collaboration and Their Benefits

PsyBench (Brave et al. 1998) first introduced physicality in CSCW, using synchronized distributed physical objects for collaborative design. Tangibles platform for collaborative work have also shown great potential for problem solving such as Urp (Underkoffler and Ishii 1999), a tangible urban planning platform in which users can collaboratively manipulate a series of physical building models and tools upon a surface. Researchers have also explored using tangible user interfaces to expand communication capabilities and help raise mutual awareness among collaborators (Brewer et al. 2007; Holmquist et al. 1999).

ReacTable (Jordà et al. 2007) later demonstrated a compelling application of tangible user interfaces for collaborative work. Multiple users could create music by simply combining and controlling tokens on a multitouch surface with dedicated function such as producing and filtering audio, adjusting parameters like volume, pitch or effects.

Hornecker and Buur (2006) identified several elements showing how social interaction is facilitated by tangible interaction systems. Our knowledge of the physical world and the interactions performed everyday with it facilitate engagement and contribution. Manual interaction with objects is easily observed and understood, advantaging group coordination and awareness. Schneider et al. (2016) demonstrated the benefits of shared physical grounding for effective collaboration using shared visual attention.

2.3 Active Tangibles for Remote Collaboration

While TUIs present clear advantages for colocated collaboration, technical limitations prevent them from supporting remote collaboration. Indeed, traditional TUIs are limited by the one-way mapping between digital and physical worlds. Manipulations of physical objects affect the digital world, yet changes in the digital world cannot be applied to the physical world. This results in mismatching states between the controls and the content (Ishii et al. 2012). To overcome this limitation, researchers have explored ways to associate computation and actuation using various technological approaches.

Many have used arrays of electromagnets to move passive tangibles on tabletops surfaces (Pangaro et al. 2002; Patten and Ishii 2007; Underkoffler and Ishii 1999; Weiss et al. 2010), while others have relied on electrostatic induction to actuate lightweight objects on a tabletop surface (Amano and Yamamoto 2012). These allow to move tangibles synchronously with the digital world, interaction is inherently limited to the surface. Others have sought to provide mobility to tangibles directly, using wheeled robots (Kojima et al. 2006; Le Goc 2016; Mi and Sugimoto 2011; Patten Studio 2014; Pedersen and Hornbæk 2011; Rosenfeld et al. 2004) or vibration (Nowacka et al. 2013). Others have explored how shape displays can allow multiple platforms to be synchronized remotely to better support collaboration (Leithinger et al. 2014; Poupyrev et al. 2007).

Leveraging this mobility, interfaces can then be synchronized in the distance to support remote interaction and collaboration. Remote Active Tangible Interactions (Richter et al. 2007) observed how active TUIs increase users' sensation of social presence and overall experience for remote collaboration compared to tradition GUIs, while others have investigated haptic feedback (He et al. 2017), collaborative planning (Riedenklau et al. 2012) or games (Mueller et al. 2006).

2.4 Tangible Augmented Reality

In parallel, the research community has looked at combining tangible user interfaces and augmented reality. Indeed, while TUIs often suffer from the lack of detailed information, AR allows to overcome this limitation by overlaying graphical context in-situ. The combination of these two technologies have led to promising platforms (Billinghurst et al. 2001; Kato et al. 2001; Kasahara et al. 2013; Lee et al. 2004; Zhou et al. 2004) where users can leverage tangible input with an augmented reality display or output. As the user manipulates tangibles in the physical world, the virtual object associated to it moves accordingly. The graphical representation of the virtual object, displayed using projection or see-through devices, is consequently always matching its tangible counterpart.

Others used AR to augment TUI in educational applications to provide additional information, explanations and simulation results on physical objects (Furió et al. 2017; Roo et al. 2017; Roo and Hachet 2017). Collaboration with virtual agents in AR is improved when using tangibles rather than holograms, as participants felt more co-presence and perceived the virtual collaborator as more physical, leading to an overall better user experience (Lee et al. 2018).

2.5 Motivations and Approach

While we have seen numerous investigations aiming at better understanding and improving collaboration in co-located contexts using tangible user interfaces or augmented reality, fewer have explore the challenges of remote collaboration and to the best of our knowledge, no existing platform or tools for remote collaboration support synchronized physical manipulations and interactions. Yet, we argue that this physical grounding will help allow for better collaboration over the distance by fostering the rich and diverse multi-modal communication channels used in collaborative contexts, in particular for creative tasks.

In order to further foster remote collaboration, we propose a platform based on active tangible interfaces (Zooids) coupled with augmented reality using head mounted see-through display (Microsoft Hololens) to display in-situ contextual information and enable better communication capabilities for richer collaboration. As it is generally agreed that good collaboration requires effective collaboration, we argue that the combination of synchronized physical grounding and in-situ contextual information better supports collaboration by providing more communication modalities and multiplying non-verbal opportunities such as gaze, hand gestures, manipulations and other demonstrations. Building on previous research, we believe this platform can provide a better remote collaboration experience using active tangibles as it increases the feeling more co-presence, and makes the collaborator perceived more physical.

Eventually, the goal of this research is to explore and better understand the benefits for tangible manipulation in remote collaboration, especially in the context of creative tasks such as tangible thinking. Thus, the resulting research question is: How can the combination of physicality and AR can improve remote collaboration, and what are the reasons for this improvement?

3 Implementation

We describe here the implementation of our platform combining AR and active tangibles to provide users with rich communication channels, physical grounding and contextual information. Each runs on identical hardware setups comprising a Zooid platform, a Hololens display and a Leap Motion tracker, described below.

3.1 User Tracking

We used a Leap Motion tracker (Leap Motion Inc 2018) to capture and monitor users' hands movements and manipulations. We placed it facing down above the table (see Fig. 1), not to interfere with the movement of the Zooids. Continuous real time hand tracking is necessary for several reason. First, it is essential to be able to see one's manipulations and hand gestures allow rich and expressive communication between remote collaborators. Indeed, communication is often supported, augmented and emphasized with hand gestures. Non-verbal communication also relies on mutual understanding of each other's action, and of particular importance here, objects manipulations. While currently available collaborative platforms provide only limited communication channels, faithful representations of the collaborators' hands in real time provide meaningful contextual information to understand, foresee and make sense of the collaborators' action.

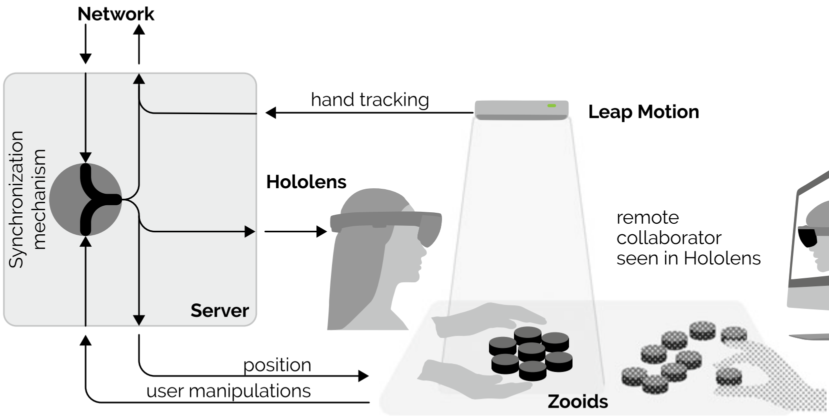


Fig. 1 The system used is composed of two identical platforms located in remote locations. Each is composed of a Hololens for the augmented reality display, a Zooid platform for the physical grounding and a Leap Motion for the hand tracking

Furthermore, accurate hand tracking also allows the system to understand users gestures, creating opportunities for a wide range of interactions with the remote collaborators as well as with the system. It also enables users to interact with holograms and other graphical representations using gestures (e.g. pinches or swipes). Regarding verbal communication, we rely on well-known tools such as Skype that package all the necessary functionalities for video conference. A computer facing the user captures video and audio flows while displaying the remote participant (see Fig. 1 right).

3.2 Augmented Reality

We used Microsoft Hololens (Microsoft Corporation 2019) as a head mounted see-through display to create a augmented reality environment, providing contextual information in-situ. Each Hololens display connects wirelessly using Wifi to receive necessary information regarding hand tracking and Zooids' positions. Each Zooid is represented in the virtual world by a hologram displayed on top of the associated Zooid.

To avoid further instrumenting the user with additional tracking hardware, we implemented a calibration sequence to register the holograms with their physical counterparts. We leverage Hololens inside-out tracking of the surrounding environment to provide a 'gaze' vector that can be made visible to the user. We then need to retrieve, from the Hololens' point of view, the position and angle of the plane of which Zooids operate and the positions of two of these Zooids. The first is obtained by reading Hololens' orientation data while setting it down on the tabletop where

Zooids roam on. We can then reconstruct the plane of the tabletop. To then continue the process, the user taps a dedicated calibration Zooid. This signals the Hololens to construct the plane based on the gathered orientation data, and sends two Zooids in preset locations.

As the gaze is displayed in the Hololens by a pointing line, the user is then asked to look at both Zooids one after the other, while grasping them. As the system is informed of which Zooid is being manipulated thanks to the touch information from Zooids, the position of each robot is known in both referentials. With the tabletop reference plane and the position of two Zooids on it, it is then trivial to compute the transformation between the Zooids' and the Hololens' referentials in order to have them match perfectly, and align both the virtual and physical worlds.

3.3 Active Tangibles

We created synchronized tangible user interfaces using Zooids, an open-source platform composed of collections of custom-designed wheeled micro robots each 2.6 cm in diameter capable of handling display and detecting user input. Each communicates wirelessly with a radio base-station, and locates itself in space using the structured light patterns from a high-speed DLP projector (Le Goc 2016). We use an updated version of the Zooids open source platform (see Fig. 2). Original Zooids are blinded by the infrared illumination commonly used by active tracking devices - here the Leap Motion for hand tracking and the Hololens headset. We modified the hardware with photodiodes (Vishay Semiconductors VEMD5510CF) capable of blocking parasitic

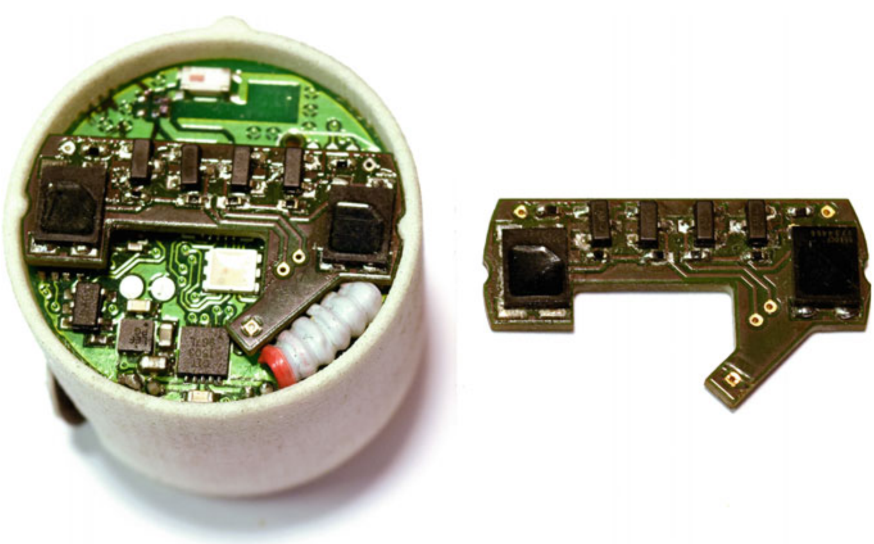


Fig. 2 A Zooid with updated hardware, using light sensor resilient to infrared illumination

infrared light (with wavelengths above 750 nm) without hindering the structured light patterns Zooids use for self-localization.

3.4 Remote Synchronization and Communication

The synchronization of the two distant stations relies on a master-slave architecture communicating with UDP packets via local network. The two platforms exchange all the necessary data 60 times per seconds to ensure a low latency user experience, without impairing communication between both parties. At the center of the communication is the native Zooid data structure. Containing positions, orientation, display colors and user interactions, it is used to convey all the necessary information to create synchronized interactive applications based on spatial manipulations. It is serialized using JSON formatting and sent to the remote station through UDP sockets. While the hand tracking data is used locally to detect interactions with holograms, it is also serialized and concatenated at the end of this packet to be sent to the distant station and displayed as holograms of the collaborators' hands. The system synchronizes the Zooids data sequentially to avoid conflicts and inconsistencies. It updates the current Zooids' statuses upon reception from the master, and with it the associated virtual objects. The interaction status of each Zooid allows the system to prioritize commands. If the user is manipulating the Zooid while the master is commanding it to move to a different location, the user manipulation is favored and sent to the master. In case of conflicting manipulations, i.e. both collaborators are manipulating the same object, the system prioritizes the first grasped Zooid information (Fig. 1).

4 Design Considerations

Effective collaboration is highly reliant on meaningful communication. In co-located cases, collaborators can use unrestricted communications modalities, verbal as well as non-verbal, to exchange and share with each other. Collaboration relies on a variety of subtle cues, turn-taking and other social constructs, including respectful and equitable interaction, in order to maintain mutual understanding between collaborators. However, while the same collaboration rules apply in remote contexts, limited communication modalities impair the clarity of the communication, creating numerous opportunities for misunderstandings and conflicts. We discuss here important aspects that have to be carefully considered when designing remote collaboration platforms involving AR and TUIs, to best entail mechanisms to prevent or at least raise everyone's awareness of the collaborators actions.

4.1 Multiplying Communication Channels

Existing remote collaboration tools and platforms only provide limited communication capabilities. Yet, strong collaboration depends on intricate communication that often spans over multiple modalities, verbal as well as non-verbal, explicit and implicit. Providing as many channels as possible allows to best leverage our natural communication capabilities. For instance, the way collaborators position themselves with respect to each in space along with their body language, the layout of the objects around them, the in-hand gestures or tone of voice while conversing are all examples of diversity in communication during collaborative tasks. They can help clarify ownership over objects and indicate limits between personal and shared spaces.

We recommend carefully identifying the critical communication channels when designing remote collaboration platforms, and multiplying these channels to provide rich and exhaustive contextual information for each collaborators. While voice and video are nowadays standard for any conference call and remote meeting, other communication channels that rely on body movement such as hand gestures remain neglected. Yet, these are powerful assets for effective collaboration, communication and demonstration complementing verbal explanation.

4.2 Preventing Concurrent Manipulations

Concurrent manipulations happen when several collaborators attempt to interact simultaneously with the same object, whether it is physical or virtual. As one manipulates an object, actions are not taken into account by the system and the object can try to move away and escape the user's grasp. While this seems obvious and would not lead to confusion in-collocated situation, the absence of information with remote platforms can leave collaborators dubious and perplexed after such unexpected behavior.

In this case, we propose to display collaborators' hands as hologram grasping the physical Zooid (see Fig. 3, while also changing its color. In case of concurrent manipulation, the hologram of the grasping hand with the Zooid separates from the physical Zooid while it provides haptic feedback to indicate the conflict. The duplicated communication channels intervene at different stages of the manipulation, allowing to preserve collaborators' awareness. Combined with the audio and video feedback, these help disambiguate potential conflicts. The platform must be as responsive as possible to ensure coherence and consistency between all the communication channels to prevent creating confusion and misunderstandings between collaborators.

5 Applications

We present here two scenarios that illustrate usage of platforms combining AR and active tangibles to enhance the collaboration experience in remote contexts.

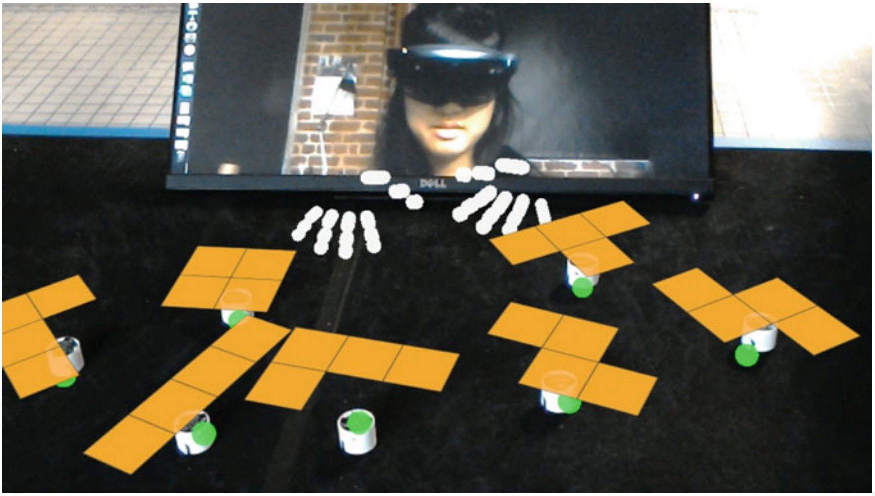


Fig. 3 View through the Hololens. Virtual objects are displayed as holograms on top of Zooids, which act as handles for manipulating digital content. The hands of the remote collaborator are displayed in relation with the physical world to display hand manipulations as well as non-verbal communication such as hand gestures

5.1 Lego Serious Play

Lego Serious Play is a methodology where LEGO bricks are used to provide with means to build physical representations of concepts, ideas and stories (see Fig. 4 Left). Fostering creative thinking and imagination, LEGO assemblies bolster communication by helping express and visualize ideas that can be difficult to express only with words, while encouraging discussions and knowledge sharing. These physical artifacts are also valuable for collaboration, as they help establish a common ground among those involved in the design and in externalizing or supporting what a designer is relaying verbally to others (Buxton 2009; Clark et al. 1991).

Using Zooids in combination with passive magnetic building blocks (Zhao et al. 2017) (see Fig. 4 Right), collaborators can assemble complex objects from a distance. As they work in remote locations, their platforms are synchronized to produce the exact same assemblies. The system can support two different modes of operation: real-time or differed assemblies. In real-time, all assemblies are synchronized at every given moment. As one starts putting pieces together, the Zooids on the other side react and move the necessary blocks to replicate the same assembly in real-time. As a result, all collaborators contribute to the same assembly. They can discuss and exchange to agree on the design, explore possibilities as if they were working in the same location, and each of their manipulations can be seen by all as holographic projections. The collaborators can also choose to first explore their own design using the differed mode. As each work on their own object, they can then sequentially

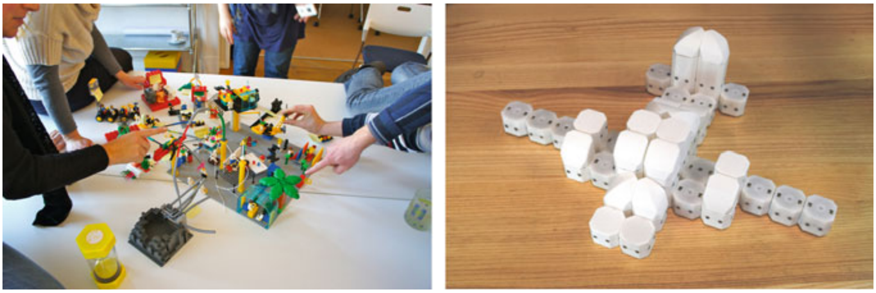


Fig. 4 (Left) A session of Lego serious play for design thinking. An example of a robotic assembly of a complex proxy object using Zooids (Right)

present their idea. A replica of the model is assembled by the Zooids to provide the best context for each presentation.

In both cases, the resulting assembly serves as reference and thus providing tangible grounding for all explanations, manipulations and other gestures. While these physical proxies provide only limited resolution and fidelity due to the size of the blocks, they can be completed with graphical overlay in augmented reality. The physicality of the platform also allows collaborators to improvise manipulations and interactions that do not need to be taken into account by the system. Using sticky notes or containers to annotate and organize objects are powerful and fruitful practices in design thinking activities. These can easily be communicated using this platform through video flow to empower all the collaborators with one's observations and insights.

5.2 Remote Collaborative Learning

Extensive investigations have demonstrated the potential of AR technologies for assistance, demonstration and learning (Henderson and Feiner 2011; Oda et al. 2015; Elvezio et al. 2017). The benefits of physical manipulation for education have long been established by seminal work from Piaget, Montessori or Fröbel, and more recent work has shown how TUIs can help support children cognitive processes (Africano et al. 2004; Antle 2007). Using this platform, children with special needs, limited mobility or studying at home, from around the world, can have access to group activities. While currently collaborative technologies already support education over a distance, previous research (Xie et al. 2008) indicates that TUIs improve engagement and enjoyment with tangible educational platforms.

The enhanced communication capabilities of this platform, combining AR and active tangibles, can help bridge the gap created by distance. Teachers and instructors can lecture STEM concepts leveraging the dynamic nature of Zooids, such as haptic feedback generated by active tangible (Özgür et al. 2017). Children can show and guide each other while working on group assignments, directly from their room, where Zooids would mingle among the furniture. The interactive physical grounding allows children to explore and experiment while preserving the comfort and confidence of known environments. We believe this can empower and encourage children to better learn and collaborate in remote contexts.

6 Limitations and Future Work

Several technical limitations appeared as we developed this platform. Firstly, the field of view of the Hololens AR display does not exceed 30°. This only allows to display only a very small portion of the collaboration space, thus forcing users to constantly tilt their head to capture different parts of the virtual collaboration space. This also increases users' cognitive load, as they need to explore and memorize the parts of the virtual world not represented by the AR display. To the best of our knowledge, no existing AR technology currently support wide field of view. Alternative technologies including projection or virtual reality headsets with see-through capabilities could potentially improve the user experience.

Furthermore, the hand tracking space conditioned by the limitations of the Leap Motion drastically restricts the collaboration space. Indeed, the typical tracking area the Leap Motion is capable of monitoring is confined to an area of about 25 cm $\times$ 15 cm $\times$ 25 cm. This does not allow to capture hand manipulation on a usual working area such as a desk. Additionally, capturing the hands from above makes it more difficult for the tracking algorithms, resulting in less precise and stable hand tracking. Another approach could use a motion capture system to monitor users movements across a significantly larger area, though at the cost of expressiveness and details such as hand gestures and other fine manipulations.

Conducting evaluations is necessary to gather additional insights and further develop the design implications for remote collaboration platforms. Not only will this allow better observations of collaborative behaviors, quantitative and qualitative analyses of collaborative tasks performed by groups will provide a deeper understanding of the impact of physicality on remote collaboration when combined with augmented reality.

The limited expressiveness of Zooids, our active tangible platform, represent a important limitation for the system presented here. The single puck-like form factor limits their usage to simple "handles" for virtual artifacts displayed in AR, rather than rich physical proxies that would leverage the wide range of our senses and capabilities. Using Zooids to create assemblies of magnetic building blocks (Zhao et al. 2017), either manually or automatically, is a first step towards enabling rich proxy objects to better support design thinking and creative tasks. Yet, it remains very challenging to create intricate and interactive objects on demand.

New opportunities emerge from these new interactive capabilities for remote collaboration. By leveraging the computation behind this dynamic tangibility that both store (i.e., keeping track of history) and recreate the artifacts' movements, we can collectively revisit our designs. When engaging in exploratory and iterative processes, we can "undo" the designs and "branch" into different alternatives in a tangible manner, thereby taking advantage of the expressivity of such dynamic physical platforms.

7 Conclusion

We presented early explorations aiming at combining augmented reality with active tangibles to better enable remote collaboration. We introduced our telepresence system to bridge the distance between collaborators and bring collaboration spaces closer together. We hope to increase the sense of co-presence and the ability of participants to communicate naturally using gaze, gesture, posture, and other body-language cues.

We contribute the first system implementation combining Zooids, a tabletop swarm interface with many robotic tangibles with Microsoft Hololens, a head-mounted see-through display and online videoconferencing tools to create a remote collaboration platform providing synchronized physical grounding and rich contextual information. We presented new challenges from the interaction design perspective, and proposed several application scenario depicting how such platform can enrich remote collaboration, in particular in creative contexts such as design thinking. Finally, we discussed the limitations of our approach and opened new doors for future work. We hope that this work will encourage researchers and designers to further explore and study remote collaboration supported by tangible interfaces, to bridge the distance between collaborators and allow for more effective work from a distance.

Acknowledgements This work is supported in part by the Hasso Plattner Design Thinking Research Program.

References

Africano, D., Berg, S., Lindbergh, K., Lundholm, P., Nilbrink, F., & Persson, A. (2004). Designing tangible interfaces for children's collaboration. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems, CHI EA '04* (pp. 853–868). ACM, New York, NY, USA. https://doi.org/10.1145/985921.985945.

Amano, K., & Yamamoto, A. (2012). Tangible interactions on a flat panel display using actuated paper sheets. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, ITS '12* (pp. 351–354). ACM, New York, NY, USA. https://doi.org/10.1145/2396636.2396698.

Antle, A. N. (2007). The cti framework: Informing the design of tangible systems for children. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07* (pp. 195–202). ACM, New York, NY, USA. https://doi.org/10.1145/1226969.1227010.

Billinghurst, M., Kato, H., & Poupyrev, I. (2001). The magicbook–Moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications*, *21*(3), 6–8. https://doi.org/10.1109/38.920621.

Brave, S., Ishii, H., & Dahley, A. (1998). Tangible interfaces for remote collaboration and communication. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative*

Work, CSCW '98 (pp. 169–178). ACM, New York, NY, USA. https://doi.org/10.1145/289444.289491.

Brewer, J., Williams, A., & Dourish, P. (2007). A handle on what's going on: Combining tangible interfaces and ambient displays for collaborative groups. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07* (pp. 3–10). ACM, New York, NY, USA. https://doi.org/10.1145/1226969.1226971.

Buxton, B. (2009). Mediaspace–meaningspace–meetingspace. In *Media space 20+ years of mediated life* (pp. 217–231). Springer.

Clark, H. H., Brennan, S. E., et al. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*, 127–149.

Elvezio, C., Sukan, M., Oda, O., Feiner, S., & Tversky, B. (2017). Remote collaboration in ar and vr using virtual replicas. In *ACM SIGGRAPH 2017 VR Village* (p. 13). ACM.

Everitt, K. M., Klemmer, S. R., Lee, R., & Landay, J. A. (2003). Two worlds apart: Bridging the gap between physical and virtual media for distributed design collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 553–560). ACM

Furió, D., Fleck, S., Bousquet, B., Guillet, J. P., Canioni, L., & Hachet, M. (2017). Hobit: Hybrid optical bench for innovative teaching. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17* (pp. 949–959). ACM, New York, NY, USA. https://doi.org/10.1145/3025453.3025789.

He, Z., Zhu, F., & Perlin, K. (2017) Physhare: Sharing physical interaction in virtual reality. In *Adjunct Publication of the 30th Annual ACM Symposium on User Interface Software and Technology* (pp. 17–19). ACM

Henderson, S., & Feiner, S. (2011). Exploring the benefits of augmented reality documentation for maintenance and repair. *IEEE Transactions on Visualization and Computer Graphics*, *17*(10), 1355–1368.

Holmquist, L. E., Falk, J., & Wigström, J. (1999). Supporting group collaboration with interpersonal awareness devices. *Personal Technologies*, *3*(1), 13–21. https://doi.org/10.1007/BF01305316.

Hornecker, E., & Buur, J. (2006). Getting a grip on tangible interaction: A framework on physical space and social interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '06* (pp. 437–446). ACM, New York, NY, USA. https://doi.org/10.1145/1124772.1124838.

Ishii, H., & Kobayashi, M. (1992). Clearboard: A seamless medium for shared drawing and conversation with eye contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 525–532). ACM.

Ishii, H., Lakatos, D., Bonanni, L., & Labrune, J. B. (2012). Radical atoms: Beyond tangible bits, toward transformable materials. *Interactions*, *19*(1), 38–51. https://doi.org/10.1145/2065327.2065337.

Ishii, H., & Miyake, N. (1991). Toward an open shared workspace: Computer and video fusion approach of teamworkstation. *Communications of the ACM*, *34*(12), 37–50.

Jordà, S., Geiger, G., Alonso, M., & Kaltenbrunner, M. (2007). The reactable: Exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07* (pp. 139–146). ACM, New York, NY, USA. https://doi.org/10.1145/1226969.1226998.

Kasahara, S., Niiyama, R., Heun, V., & Ishii, H. (2013). Extouch: Spatially-aware embodied manipulation of actuated objects mediated by augmented reality. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, TEI '13* (pp. 223–228). ACM, New York, NY, USA. https://doi.org/10.1145/2460625.2460661.

Kato, H., Billinghurst, M., Poupyrev, I., Tetsutani, N., & Tachibana, K. (2001). Tangible augmented reality for human computer interaction. In *Proceedings of Nicograph 2001*.

Kojima, M., Sugimoto, M., Nakamura, A., Tomita, M., Nii, H., & Inami, M. (2006). Augmented coliseum: An augmented game environment with small vehicles. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)* (pp. 6 pp). https://doi.org/10.1109/TABLETOP.2006.3.

Le Goc, M., Kim, L. H., Parsaei, A., Fekete, J. D., Dragicevic, P., & Follmer, S. (2016). Zooids: Building blocks for swarm user interfaces. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16* (pp. 97–109). ACM, New York, NY, USA (2016). https://doi.org/10.1145/2984511.2984547.

Leap Motion Inc. (2018). Leap motion. Online. https://www.leapmotion.com/.

Lee, G. A., Nelles, C., Billinghurst, M., & Kim, G. J. (2004). Immersive authoring of tangible augmented reality applications. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality* (pp. 172–181). IEEE Computer Society.

Lee, M., Norouzi, N., Bruder, G., Wisniewski, P. J., & Welch, G. F. (2018). The physical-virtual table: exploring the effects of a virtual human's physical influence on social interaction. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology* (p. 25). ACM.

Leithinger, D., Follmer, S., Olwal, A., & Ishii, H. (2014). Physical telepresence: Shape capture and display for embodied, computer-mediated remote collaboration. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology, UIST '14* (pp. 461–470). ACM, New York, NY, USA. https://doi.org/10.1145/2642918.2647377.

Mi, H., & Sugimoto, M. (2011). Hats: Interact using height-adjustable tangibles in tabletop interfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces, ITS '11* (pp. 71–74). ACM, New York, NY, USA. https://doi.org/10.1145/2076354.2076368.

Microsoft Corporation. (2019). Microsoft hololens | the leader in mixed reality technology. Online. https://www.microsoft.com/en-us/hololens.

Mueller, F. F., Cole, L., O'Brien, S., & Walmink, W. (2006). Airhockey over a distance: A networked physical game to support social interactions. In *Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, ACE '06*. ACM, New York, NY, USA. https://doi.org/10.1145/1178823.1178906.

Nowacka, D., Ladha, K., Hammerla, N. Y., Jackson, D., Ladha, C., Rukzio, E., & Olivier, P. (2013). Touchbugs: Actuated tangibles on multi-touch tables. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13* (pp. 759–762). ACM, New York, NY, USA. https://doi.org/10.1145/2470654.2470761.

Oda, O., Elvezio, C., Sukan, M., Feiner, S., & Tversky, B. (2015). Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 405–415). ACM.

Özgür, A., Johal, W., Mondada, F., & Dillenbourg, P. (2017). Windfield: Demonstrating wind meteorology with handheld haptic robots. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 48–49). ACM.

Pangaro, G., Maynes-Aminzade, D., & Ishii, H. (2002). The actuated workbench: Computer-controlled actuation in tabletop tangible interfaces. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology, UIST '02* (pp. 181–190). ACM, New York, NY, USA. https://doi.org/10.1145/571985.572011.

Patten, J., & Ishii, H. (2007). Mechanical constraints as computational constraints in tabletop tangible interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '07* (pp. 809–818). ACM, New York, NY, USA. https://doi.org/10.1145/1240624.1240746.

Patten Studio. (2014). Thumbles: Expanding the vocabulary of design. Online. http://www.pattenstudio.com/works/thumbles/.

Pedersen, E. W., & Hornbæk, K. (2011). Tangible bots: Interaction with active tangibles in tabletop interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11* (pp. 2975–2984). ACM, New York, NY, USA. https://doi.org/10.1145/1978942.1979384.

Poupyrev, I., Nashida, T., & Okabe, M. (2007). Actuation and tangible user interfaces: The vaucanson duck, robots, and shape displays. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07* (pp. 205–212). ACM, New York, NY, USA. https://doi.org/10.1145/1226969.1227012.

Richter, J., Thomas, B. H., Sugimoto, M., & Inami, M. (2007). Remote active tangible interactions. In *Proceedings of the 1st International Conference on Tangible and Embedded Interaction, TEI '07* (pp. 39–42). ACM, New York, NY, USA. https://doi.org/10.1145/1226969.1226977.

Riedenklau, E., Hermann, T., & Ritter, H. (2012). An integrated multi-modal actuated tangible user interface for distributed collaborative planning. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction, TEI '12* (pp. 169–174). ACM, New York, NY, USA. https://doi.org/10.1145/2148131.2148167.

Roo, J. S., Gervais, R., Frey, J., & Hachet, M. (2017). Inner garden: Connecting inner states to a mixed reality sandbox for mindfulness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17* (pp. 1459–1470). ACM, New York, NY, USA. https://doi.org/10.1145/3025453.3025743.

Roo, J. S., & Hachet, M. (2017). One reality: Augmenting how the physical world is experienced by combining multiple mixed reality modalities. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST '17* (pp. 787–795). ACM, New York, NY, USA. https://doi.org/10.1145/3126594.3126638.

Rosenfeld, D., Zawadzki, M., Sudol, J., & Perlin, K. (2004). Physical objects as bidirectional user interface elements. *IEEE Computer Graphics and Applications*, *24*(1), 44–49. https://doi.org/10.1109/MCG.2004.1255808.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *23*(6), 39.

Tang, J. C., & Minneman, S. (1991). Videowhiteboard: Video shadows to support remote collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 315–322). ACM.

Underkoffler, J., & Ishii, H. (1999). Urp: A luminous-tangible workbench for urban planning and design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '99* (pp. 386–393). ACM, New York, NY, USA. https://doi.org/10.1145/302979.303114.

Weiss, M., Schwarz, F., Jakubowski, S., & Borchers, J. (2010). Madgets: Actuating widgets on interactive tabletops. In *Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology, UIST '10* (pp. 293–302). ACM, New York, NY, USA. https://doi.org/10.1145/1866029.1866075.

Wilson, A. D., & Robbins, D. C. (2007). *Playtogether: Playing games across multiple interactive tabletops*. In IUI Workshop on Tangible Play: Research and Design for Tangible and Tabletop Games.

Xie, L., Antle, A. N., & Motamedi, N. (2008). Are tangibles more fun? Comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction* (pp. 191–198). ACM.

Yarosh, S., Tang, A., Mokashi, S., & Abowd, G. D. (2013). Almost touching: parent-child remote communication using the sharetable system. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 181–192). ACM.

Zhao, Y., Kim, L.H., Wang, Y., Le Goc, M., & Follmer, S. (2017). Robotic assembly of haptic proxy objects for tangible interaction and virtual reality. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces* (pp. 82–91). ACM.

Zhou, Z., Cheok, A. D., Chan, T., Pan, J. H., & Li, Y. (2004). Interactive entertainment systems using tangible cubes. In *IE2004 Proceedings of the First Australian Workshop on Interactive Entertainment* (p. 19).

DT@IT Toolbox: Design Thinking Tools to Support Everyday Software Development

Franziska Dobrigkeit, Philipp Pajak, Danielly de Paula and Matthias Uflacker

Abstract Recent evidence suggests that design thinking can contribute to agile software development by increasing attention towards user needs as well as collaboration in interdisciplinary teams. However, there is a lack of understanding and support on how to facilitate the integration of design thinking activities into the daily work of agile development teams. Our work extends existing research on integrating design thinking with agile software development methodologies by developing and validating a toolbox that software teams can use in their everyday work. We present the DT@IT Toolbox, a collection of design thinking methods targeted at design thinking novices that aims to support everyday software development activities. The toolbox was evaluated with a team from an SME based in Germany over a period of 12 weeks. As a result, participants reported that using the DT@IT Toolbox led to a better communication within the team, enhanced problem-solving skills, increased empathy towards users and led to a better understanding of the users needs.

1 Introduction

There have been serious attempts to integrate more design efforts to agile development to increase attention towards the user and enhance the team's problem-solving skills. In particular, Valencia et al. (2013) claim that identifying the variety of roles that designers can fulfil in companies is pivotal to supporting the strategic utilisation of design, and strengthening the product development processes. However, much un-

F. Dobrigkeit (✉) · P. Pajak · D. de Paula · M. Uflacker
Hasso Plattner Institute for Digital Engineering, Campus Griebnitzsee,
14482 Potsdam, Germany
e-mail: franziska.dobrigkeit@hpi.de

P. Pajak
e-mail: espspinix@gmail.com

D. de Paula
e-mail: D.FERREIRAOLIVEIRADEPAULA1@nuigalway.ie

M. Uflacker
e-mail: matthias.uflacker@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_13

certainty still exists about how to optimise the relationship among interdisciplinary teams. This might happen because team members have different expectations, motivations and use a different language to describe product features. In order to overcome this situation, researchers propose integrating design thinking (DT) into agile software development (Hildenbrand and Meyer 2012; Grossman-Kahn and Rosensweig 2012; Dobrigkeit et al. 2018; Gurusamy et al. 2016; Ximenes et al. 2015). As one interviewee stated in De Paula et al. (2018): "once you teach everyone design thinking, which is fundamentally focused on the user need, we have a common language to talk about (...), so design overcomes the semantic gap where we don't share common languages". By providing a common language, design thinking can support teams to increase collaboration and interdisciplinary teamwork (Carlgren et al. 2014). Carlgren et al. (2014) mention that collaboration can promote better team dynamics in terms of increased energy, inspiration and motivation. Considering this evidence, it seems that design thinking can contribute to agile software development, not only as a pre-development activity but also as an activity supporting development. However, there is a lack of support on how to facilitate the integration of design thinking activities into the daily work of agile development teams. In light of this, the goal of this study is to support agile development teams with beneficial design thinking activities. To fulfil this aim, we identified and organised design thinking methods in a toolbox that can support software teams during everyday agile development.

By employing concurrent triangulation mixed methods design, we applied 12 DT methods over a period of 12 weeks in one industry team. The application consisted of introducing one method each week and letting the team find situations to apply it at least twice during the week. We collected qualitative and quantitative data simultaneously. In order to capture data on the participants' feelings towards the benefits of using the respective method, we conducted interviews before the application period, during and after it. Furthermore, we created a questionnaire to evaluate each method at the end of the week. Additionally, we adapted the Toronto Empathy Questionnaire to create two new questionnaires that measure empathy towards team members and towards the user. They were answered once before the application period and once after all methods had been introduced and used to measure if the application period had an effect on empathy.

To support teams in their efforts to integrate design thinking methods in everyday agile software development, this study provides a comprehensive Toolbox that software teams can use to select and apply DT methods according to their needs. We provide evidence supporting the benefits of design thinking methods used in the daily work of agile development teams. They specifically help in identifying problems, finding possible solutions, improving communication and strengthening empathy. Additionally, we provide evidence that regularly applying design thinking methods within an agile team increases the level of empathy towards the user and within the team. Overall, the main benefit of the DT@IT Toolbox is to support software teams in the challenging task of integrating design thinking into daily agile work by offering a selection of beneficial design thinking methods targeted at software teams inexperienced with design thinking.

The remainder of this article is structured as follows: Sect. 2 provides an overview of existing research on the integration and benefits of design thinking and agile for software development. Our research approach is described in Sect. 3 and an overview of the DT@IT Toolbox is presented in Sect. 4. Section 5 presents our findings, and the findings are discussed and limitations are presented in Sect. 6. Finally, Sect. 7 closes this chapter with a conclusion and future work.

2 Background and Related Work

Agile methods, such as Kanban, Scrum or XP have been recommended for software development due to their benefits in relation to reducing the development time, increasing the flexibility of the process and improving the quality of the product (Erickson et al. 2005, p. 89). In particular, agile processes are designed to be flexible in uncertain environments, with mechanisms such as feedback loops, self-organising teams and small development cycles (Schwaber and Beedle 2001). However, agile teams have a strong focus on technical aspects (Ximenes et al. 2015) rather than targeting the needs of end users (Sohaib 2018). In accordance with this aspect of agile, researchers identify a lack of empathy towards the end user (Ximenes et al. 2015), a lack of attention towards design (Dybå and Dingsøyr 2008) and problem understanding and solution finding (Lindberg et al. 2011) as limitations of agile approaches. These limitations can lead to severe consequences such as a company launching the “wrong” products. This leads to in poor market reception or necessary rework requiring extra engineering hours and investments (Verganti 1997; Griffith 2014).

There have been serious efforts to integrate design and a stronger focus on the end user into agile development, in order to overcome this restriction, e.g. UCD, UI and UX design. A common discussion with such approaches is whether to have a big design upfront, or let the designers run one or two sprints before the development, or whether to integrate them into the agile development team (Kollmann et al. 2009). Another point of discussion is how much user contact is necessary and who should be involved. While the integration of design into software development processes brings a stronger attention to design and the user to agile software development, collaboration between developers, designers, and users is often difficult. A possible explanation for that might be that all participants work in a unique context with specific language, expectations, motivation and perceptions,thus making it hard to collaborate successfully (Sonnenwald 1995). For example, different communities have different expectations when it comes to design for agile software development: designers work towards a consistent and coherent design (Ferreira et al. 2007) throughout the software, while developers expect design to be done when required true to the agile principles “working software over comprehensive documentation” and “responding to change over following a plan” (Beck et al. 2001).

In order to enhance collaboration and encourage more interdisciplinary teamwork, researchers have integrated design thinking into agile development (Lindberg et al. 2011; Hildenbrand and Meyer 2012). In agile teams that closely collaborate with non-developers (e.g. designers, managers or end users) design thinking can facilitate collaboration by providing a common process, toolset, language and mindset and allowing non-designers to be part of the conceptualisation of the product (De Paula et al. 2018). Accordingly, one of the core benefits of the design thinking approach is its human-centredness, helping to develop empathy towards users and understand their wants and need (Brown 2008; Porcini 2009; Ward et al. 2009; Clark and Smith 2008). Additionally, empathy towards co-workers and better collaboration across professional borders are an important byproduct of design thinking (Clark and Smith 2008), as the interdisciplinary approach naturally balances technical, business and human aspects (Holloway 2009). Accordingly, several authors have proposed combined process models that integrate DT into (agile) software development models. Several models use DT in an upfront phase. For example, the integrated Design Thinking and Lean Development approach (Hildenbrand and Meyer 2012), the Nordstrom Innovation Lab process (Grossman-Kahn and Rosensweig 2012), InnoDev (Dobrigkeit et al. 2018) or Converge (Ximenes et al. 2015). Additionally, researchers find value in using design thinking during everyday development. For example, the Integrated design thinking and agile framework for digital transformation (Gurusamy et al. 2016), the Human-centered Agile Workflow (Glomann 2018) and DT@XP (Sohaib 2018) regularly integrate smaller design thinking phases into agile development cycles. Similarly, InnoDev (Dobrigkeit et al. 2018) and Converge (Ximenes et al. 2015) suggest the use of design thinking in later stages of the development process in addition to running a design thinking pre-phase. InnoDev proposes running DT phases when new features are added to the scope of the product and makes use of DT-Breakouts in case of blockers. Converge suggests the use of DT-Knots in case the team wants to solve a specific problem. According to Ximenes, such knots are workshops facilitating one step of the design thinking process. These two approaches to using design thinking during development resemble the UX Toolbox, an approach to train developers in facilitating usability methods during development, thereby easing UX-experts workloads and facilitating better collaboration (Pedersen 2016).

In a case study with a software development team, Dobrigkeit and de Paula indeed found all the manifestations of design thinking the aforementioned approaches prescribe (Dobrigkeit and de Paula 2019). They conclude that knowledge and use of design thinking differs for various roles within the agile team. For example, for developers it suffices that they take part in design thinking activities and have basic tool knowledge. However, there is a lack of support for agile teams in how to acquire such tool knowledge.

To sum it up, researchers see value in integrating design thinking to a different degree into agile software development processes. Suggested processes often take a high-level approach to describing such an integration and a lot of focus is put on supporting the conceptualisation of the product in the form of pre-development activities. Suggested use of design thinking during everyday software development is researched and described in a rudimentary form. And while it was shown that the application of single design thinking methods already adds value to agile software development, it is still unclear how software developers can be supported in using design thinking tools during everyday agile work. We aim to address this problem with the DT@IT Toolbox, a collection of design thinking methods aimed at supporting software development activities and targeted at design thinking novices.

3 Methodology

This section presents the research methodology and scientific approach used to develop the DT@IT Toolbox. The phases of the research, data collection methods and best practice protocols used in this research are presented and discussed.

The goal of this study is to support agile development teams with beneficial design thinking activities. For that purpose, we wanted to identify and organise design thinking methods in a toolbox that can support software teams during everyday development. The following research questions guided this study.

RQ1: What methods from design thinking can support agile development teams in their everyday work?

RQ2: How can we organise the DT methods in a way that enables software teams to use them by themselves?

As mentioned before, authors state that empathy is one of the core elements of design thinking. Therefore, in order to gain more insights into the benefits of the selected methods and how they are correlated to empathy, we formulated our third research question.

RQ3: How is empathy towards users and/or team members affected by regularly using DT methods?

In this way, in order to reach our goal and answer our research questions, we propose to develop the DT@IT Toolbox to support software teams in selecting and using methods beneficial for agile development.

3.1 Research Design (Mixed Methods Approach)

The research was developed in the context of concurrent triangulation mixed methods design. Mixed methods research is defined as "*research in which the investigator collects and analyses data, integrates the findings, and draws inferences using both qualitative and quantitative approaches, or methods, in a single study*" (Tashakkori and Teddlie 2010). An effective strategy of inquiry in mixed methods research is concurrent triangulation design (Creswell and Plano Clark 2017). The purpose of this design is to implement quantitative and qualitative methods during the same time-frame and with equal weight in order to obtain different but complementary data on the same topic" (Morse 1991, p. 122). This design was adopted because we wanted to directly compare quantitative statistical results with qualitative findings to best understand the research problem as suggested by Creswell and Clark (2017). Figure 1 illustrates the phases carried out in this research.

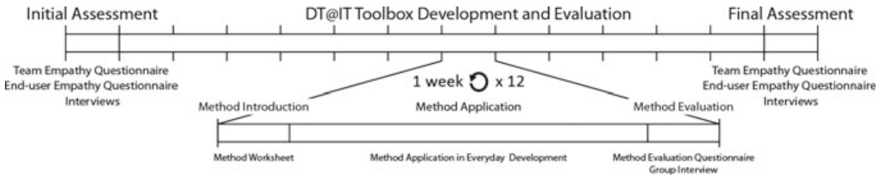


Fig. 1 Overview of how participants answered the method evaluation questionnaires

The development and evaluation of the DT@IT Toolbox was split into three phases: (i) Initial Assessment, (ii) DT@IT Toolbox Development and Evaluation, and (iii) Final Assessment. Each phase is described in the following sections.

3.2 Initial Assessment

This phase aims to understand the team's situation and problems and analyse the level of the subjects' empathy before the application of any design thinking methods. In order to do so, two questionnaires and one-to-one interviews with the subjects were carried out. The questionnaires were developed based on the Toronto Empathy Questionnaire (Spreng et al. 2009), a self-report style, uni-dimensional, 16-item questionnaire developed to assess the empathy levels of individuals. It follows a 5-point Likert scale ("never", "rarely", "sometimes", "often" and "always"). The total score of the questionnaire is calculated as the sum of all item scores (Spreng et al. 2009). We adapted the original items of the Toronto Empathy Questionnaire to suit the scenario of a software company and set the context of the items to users or team members for the two questionnaires. For instance, the sentence "When a friend starts to talk about his/her problems, I try to steer the conversation towards something else" from the Toronto Empathy Questionnaire was changed to:

- **Team Empathy Questionnaire**: "When a team member starts to talk about his/her problems, I try to find an easy way out by delegating the issue to somebody else or by proposing a quick-fix."
- **User Empathy Questionnaire**: "When an end user starts to talk about his/her problems with the software, I try to steer the conversation towards something else e.g. I believe that the problem lies with the user not the software."

In an effort to gain additional insight into the subjects' feelings about their team and process, and their opinions on empathy before the Method Application phase, we conducted focused interviews as described by Flick (2014). Focused interviews have a predefined set of questions, but in contrast to structured interviews they allow alternating or changing questions if necessary to get more details about certain experiences. The scope of focused interviews starts very broadly and narrows towards the end, in order to collect unforeseen insights as well as those that are foreseen. The interview questions were split into three sections: (i) the team and its process

(ii) empathy towards the team and (iii) empathy towards the users. Based on the information about the team and its process we could select design thinking methods to support the team, while the empathy assessment served as an initial assessment to compare our final results.

3.3 DT@IT Toolbox Development and Evaluation

The Method Application phase aims to apply 12 DT methods to a software team during a period of 12 weeks. There are many design thinking methods currently being recommended by academics and practitioners. Due to the time constraint and the team's background, we could not test all available methods and therefore had to define criteria to select the 12 DT methods. As sources for our method selection we chose method sets that are targeted at design thinking novices and created by established companies in the area of design thinking implementation, namely: the Design Thinking Prototyping Cardset (Meinel et al. 2013), IDEO Method Cards (IDEO 2003) and d.School Stanford design thinking bootleg (d.School Stanford 2018). The Design Thinking Prototyping Cardset, created at the Hasso Plattner Institute, consisting of 36 prototyping methods for design thinking novices (Jobst and Meinel 2013). The IDEO Method Cards were developed as a showcase and facilitator to inspire the design thinking process with 51 methods (IDEO 2003). The d.School Stanford design thinking bootleg (d.School Stanford 2018) is a guide with mindsets, a description of the DT process and 38 methods that aim to support design thinking implementation. When selecting 12 methods from these sources we aimed for methods that are:

- widely applicable,
- short in duration,
- easy to understand and implement,
- able to support development activities,
- comfortable for the developers.

Wide applicability is important to ensure that they can be useful in a variety of situations during the software development. The duration of the method applicability should be reasonable with day to day agile work, ideally less than 60 min. Additionally, methods should be easy to understand and be applicable by the team without special training or the support of a design thinking coach. Based on the initial interviews, we identified the following development activities as supportable by design thinking methods:

- understanding the product,
- empathizing with the team,
- empathizing with the end user,
- ideating on new features or implementation concepts,
- prototyping new features or implementations (e.g. UI, software architecture).

- giving and collecting feedback (e.g. within reviews and retrospectives, within user testing, between team members).

Finally, we decided that activities that are potentially uncomfortable to our participants, such as performing a theatre role play, or making an advertising video, should not be part of our toolbox.

The study was conducted in DESIGN-IT, which is a medium sized software development company based in Germany. From that company, a team composed of seven programmers, two designers and one team assistant were the subjects for this research. The Method Application phase lasted 12 weeks and each week followed the same structure: method introduction, method application and method evaluation. At the beginning of the week the design thinking method for the week was introduced. For that purpose, a worksheet that contains a description of the design thinking method, how to use it and some examples was developed and presented to the team. During the week the team was asked to apply the method during their daily agile work at least twice whenever they saw fit. Additionally, the team was asked to post any artefacts and results of their method application to a team chat. At the end of each week we conducted a group interview, to capture the subjects' feelings and general impressions towards the implemented method. Additionally, the method evaluation questionnaire was developed and filled out by each participant. The questionnaire captures (i) the benefits of the model (ii) usability, (iii) comprehensiveness, and (iv) the applicability of the method. The questionnaire includes eight closed format questions with differential scales and six open format questions. Thus it is a mixture of open and closed formats as described by Leung (2001). This format enabled us to collect statistically comparable results, but also enabled the participants to freely express their feelings and thoughts about the method.

3.4 Final Assessment

The Final Assessment phase aims to analyse whether our toolbox added value to the software development process and if the team members' empathy was enhanced. First, the Team Empathy Questionnaire and the User Team Questionnaire from the Initial Assessment phase were reapplied. Additionally, another round of interviews was conducted in order to capture data on the benefits and challenges of using our toolbox and a perceived change in empathy.

The quantitative data from the questionnaires was statistically evaluated by comparing the total scores from each participant before and after the teams implemented all selected design thinking methods. The total score was calculated in the same way as in the Toronto Empathy Questionnaire (Spreng et al. 2009). The questionnaires have a test-retest reliability and have already been used by other researchers for comparing empathy levels before and after certain events (Patterson et al. 2017; Hicks 2015). Furthermore, we used the Wilcoxon Signed Ranks Test to calculate the significance of our results, in the same way as described by Patterson et al. (2017).

Table 1 Overview of questionnaires and interviews with the respective research questions they address

	RQ1	RQ2	RQ3	Application
Team empathy questionnaire			x	Initial + final assessment
User empathy questionnaire			x	Initial + final assessment
Method evaluation questionnaire	x	x	x	Weekly
Interview	x	x	x	Initial + final assessment
Group interview	x	x	x	Weekly

The interview guide for the final interview round differed from the initial interview guide. Instead of a section about the team and its process, this interview guide contained a section that asked when and how often the methods were used and what problems were faced when applying them during daily work. Thus, we were able to compare qualitative and quantitative data from the Initial Assessment phase to the Final Assessment phase.

The qualitative data of this study (solo interviews during initial and final assessment as well as weekly group interviews during the evaluation phase) was analysed using coding as proposed by Flick (2014). For the evaluation of the final and initial interviews, we created different categories and coding guidelines based on patterns derived from our initial analysis. Finally, each interview was coded for each category by using these guidelines. Additionally, we analysed the group interviews to capture how often the methods were used, the challenges concerning their application and the possible application scenarios. An overview of how the implemented methods help to answer the research questions can be seen in Table 1.

4 DT@IT Toolbox

This section presents the Methods we selected to be part of our DT@IT Method Toolbox. Further information on the toolbox and the worksheets we created as part of it can be downloaded at https://bit.ly/DT4IT. As described in Sect. 3.3, we selected methods that support a wide range of applications, are widely applicable during development, easy to understand, short in application and support activities of everyday agile development, while avoiding uncomfortable activities. We selected the methods from existing method collections as also described in Sect. 3.3. Table 2 lists all methods initially selected to be part of the DT@IT method toolbox, with the activities they support in our toolbox.

Personal Inventory In this method peers share information about important items (IDEO 2003) and thus "allow the designer to see and understand the relevance of objects in a user's life from the participant's point of view to inspire design themes and insight (Hanington and Martin 2012)". In order to adjust this method for software

Table 2 List of selected methods with activities they support

Method	Activities
Personal inventory	Empathize with team/empathize with end user
Character profiles	Empathize with end user
Customer journey map	Understand product/empathize with end user
Five whys?	Communicate
Letter to grandma	Communicate/understand product
A beginner's mind	Ideate
30 Second sketch	Ideate
Powers of ten	Ideate
Empathy tools	Prototype
Paper prototype	Prototype
Feedback capture grid	Collect feedback
Five finger feedback	Collect feedback

developers we asked people (developers and end users) to share annotated pictures of their work space and describe how they work there. A similar adjustment of Personal Inventories was used by Jalleh when co-designing an air ambulance cabin with a medical flight crew (Jalleh 2013). Thus, this method provides team members with a glimpse of how everybody works and which tools are important to them. Additionally, seeing end user work spaces provides developers with an overview of how and where the end users are using their software.

Character Profiles Character Profiles, also called Personas, are fictional characters (Cross 2003), that are "based on observation of real people" (IDEO 2003). A finite number of personas can be created per project (Kumar 2012), that "in turn [represent] a group of real consumers with similar characteristics" (Miaskiewicz and Kozar 2011). Each Character Profile has a name, image (Cross 2003) and a further description about what the character likes, dislikes, its occupation and needs and goals with regard to the project (Miaskiewicz and Kozar 2011). Creating the personas is not a focus of day to day agile software development, and, as such, should be done in a separate research phase (Hanington and Martin 2012). However, working with Personas created in earlier stages enables developers to focus on their user/user group and communicate about them (Grudin and Pruitt 2002) during later development stages, thus, fostering empathy, challenging assumptions and preventing self-referential design (Miaskiewicz and Kozar 2011).

Customer Journey Map A Customer Journey Map, also known as a User Journey Map (Kumar 2012) or Journey Map (d.School Stanford 2018), visualises and describes the journey of consuming a product or service from the end users point of view (Lemon and Verhoef 2016). It details important activities and associated emotions and thus reveals insights about the strengths and weaknesses of the solution and the new opportunities to tackle (Kumar 2012). For example, a journey could start by having a need, finding a product/service to fulfil that need and end by consuming a product/service and recommending it (Rosenbaum et al. 2017, p. 2). For Agile Teams this method can help to identify usability issues within the software and surrounding services. Furthermore, a visual Customer Journey Map enables the developers to discuss and collect data around the customer experience and thus could foster product understanding and end user empathy.

Five Whys Five Whys is a simple tool, that addresses the problem of fighting symptoms without knowing the root cause and only getting short-lived results (Ohno 1988) by repeatedly asking "Why?" until one is satisfied with the answer. Thus it is "quick and easy to use [... and] leads the team to a better understanding of the issue" by helping to discover the root cause of a problem (Pojasek 2000). The method was invented by Toyoda Sakichi to explore the cause-and-effect relationships (Serrat 2017) in industrial manufacturing, but is applicable to any field. During software development it might be necessary to determine root causes, e.g. a specific customer desire might be based on a root desire that could be solved in a much more efficient way. In such cases asking, digging deeper and ensuring assumptions can lead to a better understanding of your conversation partner.

Letter to Grandma "Letter to Grandma" is a method to improve sharing and understanding of requirements and tasks (Aerssen et al. 2018). It allows a knowledge transfer that focuses on the essential aspects (Meinel et al. 2013) without omitting core functionality and benefits (Aerssen et al. 2018). Writing a letter to grandma helps to minimise knowledge sharing problems when dealing with requirements or tasks to be shared (Aerssen et al. 2018). Grandma has no prior knowledge and thus needs a description of the essentials enriched with further information that is only known to the team. Concentrating on the essentials makes sure that one sticks to the important parts, does not overwhelm the recipient with unnecessary details, but provides all the necessary knowledge with a good overview of the requirements (Aerssen et al. 2018). Because of that, we see it as a useful tool e.g. to explain epic user stories to the team or to introduce a new person to a new project.

A Beginner's Mind The general idea for this method comes from Zen Buddhism and is called "shoshin" (Suzuki 1970), which means "beginner's mind" in English. This state of mind provides an attitude of openness and a lack of preconceptions (Suzuki 1970). To apply this method one can either try to personally adopt a beginner's mind oneself or ask a real beginner - either somebody outside of the team who does not have much knowledge about the project, or a stranger on the street. Our experiences with a topic lead to assumptions, perceptions and stereotypes that we take for granted (d.School Stanford 2018). By applying A Beginner's Mind we open ourselves to unexpected discovery and innovation because we are not blinded by our prior assumptions (Belshee 2005) and have a better view of the details and the bigger picture. This method has been applied with software development by Belshee who found that knowledge sharing improved during Pair Programming, when keeping the pair in a beginner's mind by having a new partner every 90 min (Belshee 2005). Belshee states that A "Beginner's Mind is a very efficient way to solve programming problems" (Belshee 2005). Thus we expect this method to have a similar effect in our toolbox.

30 Seconds Sketch In the early ideation stage this method is used like brainstorming to generate many ideas in a short time, but additionally makes use of visualisation (Meinel et al. 2013). A sketch can be seen as a simple and low fidelity prototype - it is quick (Meinel et al. 2013) and thus easily disposable. The method works simply by creating sketches in no more than 30 seconds on paper (Meinel et al. 2013). Depending on the phase of development, one can adjust how deep to dive into the solution space (Greenberg et al. 2011). This exercise can be performed with multiple people and also iteratively by drawing multiple sketches in a row (Greenberg et al. 2011). Sketches in general are a good tool to express, develop and communicate ideas (Greenberg et al. 2011). By visualising an idea one can make it much more tangible for their peers, which enables gathering better critique and feedback as a result (Greenberg et al. 2011). As such, this method can support discussions about UI or software architecture.

Powers of Ten The method Powers of Ten is a reframing technique based on the short film of the same name from 1977 (Eames and Eames 1977). The movie shows a picnic scene in Chicago with different magnitudes of distance from 1 m up to 10^{24} meters and then down to 10^{-16} m. It takes the viewer up into the deep universe and then back into the hand of the man at the picnic until only atoms are visible. Powers of Ten is an exercise that allows changing the point of view by varying magnitudes of context for insight generation (d.School Stanford 2018). Powers of Ten makes use of exponential growth to change the point of view. This is difficult because the human brain is trained to think linearly (De Langhe et al. 2017). However, this techniques makes it possible to find the right framing for different aspects of a certain problem by seeing them from very different magnitudes - basically by zooming in and out the of current scene (d.School Stanford 2018). This method can support problem solving or insight development, e.g. thinking about different valued products from soft-drinks to houses, when designing a web shop.

Feedback Capture Grid The Feedback Capture Grid is a method to gather feedback in a structured grid (d.School Stanford 2018). It is based on four quadrants: likes, dislikes, ideas and questions labeled with easy to understand symbols that are used as categories for the feedback (d.School Stanford 2018). This method helps to be systematic about feedback and forces the user to think about different aspects (Lewrick et al. 2017). It provides an easy way to present results and makes it possible to see common themes and areas that need more development (Ritchie et al. 2016).

Five Finger Feedback Five Finger Feedback gathers differentiated information by assigning each finger a specific meaning and using those associations to give structured feedback (Brauneck et al. 1995):

Thumb	Expressing what was liked
Index Finger	What was noticed, or something to point out
Middle Finger	Expressing what was not so good, or could be done better
Ring Finger	What was emotionally interesting
Little Finger	What was missing or did not get enough attention

This method fosters constructive feedback and considers different aspects. Furthermore, it gives clarity, can improve communication in groups and allows every participant to formulate his or her statements. This method was proposed to us by one of the team members during our initial interview and we decided to include item despite it not being mentioned in our method sources (see Sect. 3).

Empathy Tools Empathy Tools is a method derived from inclusive design (Hitchcock and Taylor 2003) and used to describe physical objects and cognitive or social techniques that provide designers with a feeling and deeper understanding of users with various abilities (IDEO 2003). For example, Patricia used empathy tools as a young woman to put herself in the shoes of elderly people by using glasses to blur her vision, earplugs to hear less and uneven shoes to be forced to walk with a cane. Empathy Tools is an informal method that has no defined process to follow but is about discovering the end users abilities to get a deeper understanding of special conditions that are characteristic of the end users. Especially in the field of software development, it is crucial to see the differing abilities in using technical devices. While a team developing software has mostly very advanced technical abilities, their end users often have limited technical abilities. People with minimal or almost no technical abilities are still common, as "widespread assessment of IT literacy is just beginning to emerge" (Pérez and Murray 2010). Empathy Tools can help gain insights on such extreme users.

Paper Prototype Paper Prototyping is a method to create low fidelity prototypes based on sketches made on paper (Snyder 2003) or a whiteboard. A very basic Paper Prototype consists of sketches of each screen, a person in charge of changing the screens and a tester who is interacting with the paper version of the interface (Snyder 2003). Paper Prototypes can be used in early testing or in brainstorming meetings and sessions. A big waste in software development is spending time on unnecessary features (Frye and Inge 2013). Low fidelity prototypes, like Paper Prototyping, could help to identify problems beforehand (Snyder 2003) and is very inexpensive in terms of time and effort (Tam 2001). On paper, one can explore many different versions without spending too much time and, if necessary, also discard any idea that might not be appropriate. Compared to digital prototypes, paper has a higher flexibility in a faster and cheaper way, because it does not need any technological knowledge (Tam 2001). Additionally, prototyping on paper gives non-designers the chance to participate in the process (Moser 2012, p. 166) and thus might facilitate idea communication and decision making in the team.

5 Results

Our results show that team members liked the DT@IT Toolbox and are eager to keep on using the methods. We could further show, that an extensive DT phase using our method toolbox has improved team empathy and end user empathy. Based on the method evaluation questionnaires and group interviews we conclude that the toolbox methods were well understood by most of the participants. Table 3 provides an overview of the understandability ratings for all methods. In cases where team members were uncertain how to perform a method, other team members shared their

Table 3 Methods and their understandability (Scale 5-easy to 1-hard)

Method	How easy was it to understand the method
Five whys	4.750
Feedback capture grid	4.714
Empathy tools	4.625
Personal inventory	4.500
Letter to grandma	4.429
Paper prototype	4.429
Five finger feedback	4.286
30 Second sketch	4.250
Powers of ten	3.800
Customer journey map	3.750
A beginner's mind	3.714
Character profiles	3.333

knowledge and thus the team was able to apply all methods based on the provided worksheet without further schooling or a design thinking coach. We observed this e.g. for the Character Profiles method. Though the participants were not sure how to apply the method at first, they started exchanging examples and ideas about the application of the method and finally were able to achieve useful results. Half of the participants gave the presentation of the methods the highest possible grade - and if one also considers the second highest grade it is already above 80%. The majority of the participants who were not satisfied with the presentation, expected more examples or a better explanation on result presentation techniques. On the other hand, there were also people who believed that the description was too long. So clearly there exists a discrepancy between team members that can fully and intuitively understand the methods, and others that have difficulties.

Even though most methods were perceived as useful and participants had no problems understanding them, the pickup rate was rather low after the method evaluation phase. While the participants can imagine many possible applications for the methods, they still lack the ability to spot these moments in daily work and would like to be supported in this regard. Participants have reported that methods have been used primarily to solve the given tasks during the method evaluation phase. As summarised in Table 4, five attendees have also used selected methods for further daily tasks, but only three collaborators have done this consciously. The other two teammates have reported that they applied the methods, but as a participant explains "maybe you apply something but you don't notice [it ...] like A Beginner's Mind [...] you really don't know something and you are applying [the method] without noticing". On the contrary, three participants have reported that they have not been using the methods besides the evaluation phase. One employee explains that he had "[several] opportunities to use them [... but he] didn't use them" - he further adds

Table 4 Method usage after evaluation phase

Usage/participant	1	2	3	4	5	6	7	8	∑
No usage			X			X	X		3
Unconscious usage				X	X				2
Conscious usage	X	X						X	3

Table 5 Further method usage in the future reported by participants

Further usage/participant	1	2	3	4	5	6	7	8	∑
Yes					X	X			2
Yes, some methods	X	X	X	X			X	X	6

that “[he doesn’t] know exactly why, because [he likes] the methods and [he knows] they are useful - and [he knows] if you apply them the result is better.”

However, all participants reported they would like to use at least some of the methods in the future, as summarised in Table 5. While just two collaborators stated that they would like to use the methods in general, six attendees claimed that they would rather use a selection of the methods, mostly the ones that they have claimed to be their favourites. There was no participant that did not want to keep on using the methods.

5.1 Method Evaluation

An overview of the evaluation can be found in Fig. 2. Each circle represents how participants rated the questions (columns) of each applied method (rows). Each pie chart shows the distribution of answers on the scale. As expected, all the methods have different weak and strong points. No method proved to be beneficial in every aspects and the results show that different problems should be solved by different methods.

For each of the methods we collected the benefits as reported by our participants. From the observations that we made during the method evaluation phase, we have reached the conclusion that the ability to solve and find problems in day-to-day work, as well as better communication, support the work of developers in a human-centred way. The reported benefits are represented in Table 6.

Additionally, we collected possible applications during agile development for each method based on the method evaluation questionnaire and the group interviews, as depicted in Table 7. Most methods can be used regularly, as participants mentioned, e.g. they can be used for every new feature, or when stuck. This was not the case with Personal Inventory. Participants had doubts as to whether this method could be applied repeatedly.

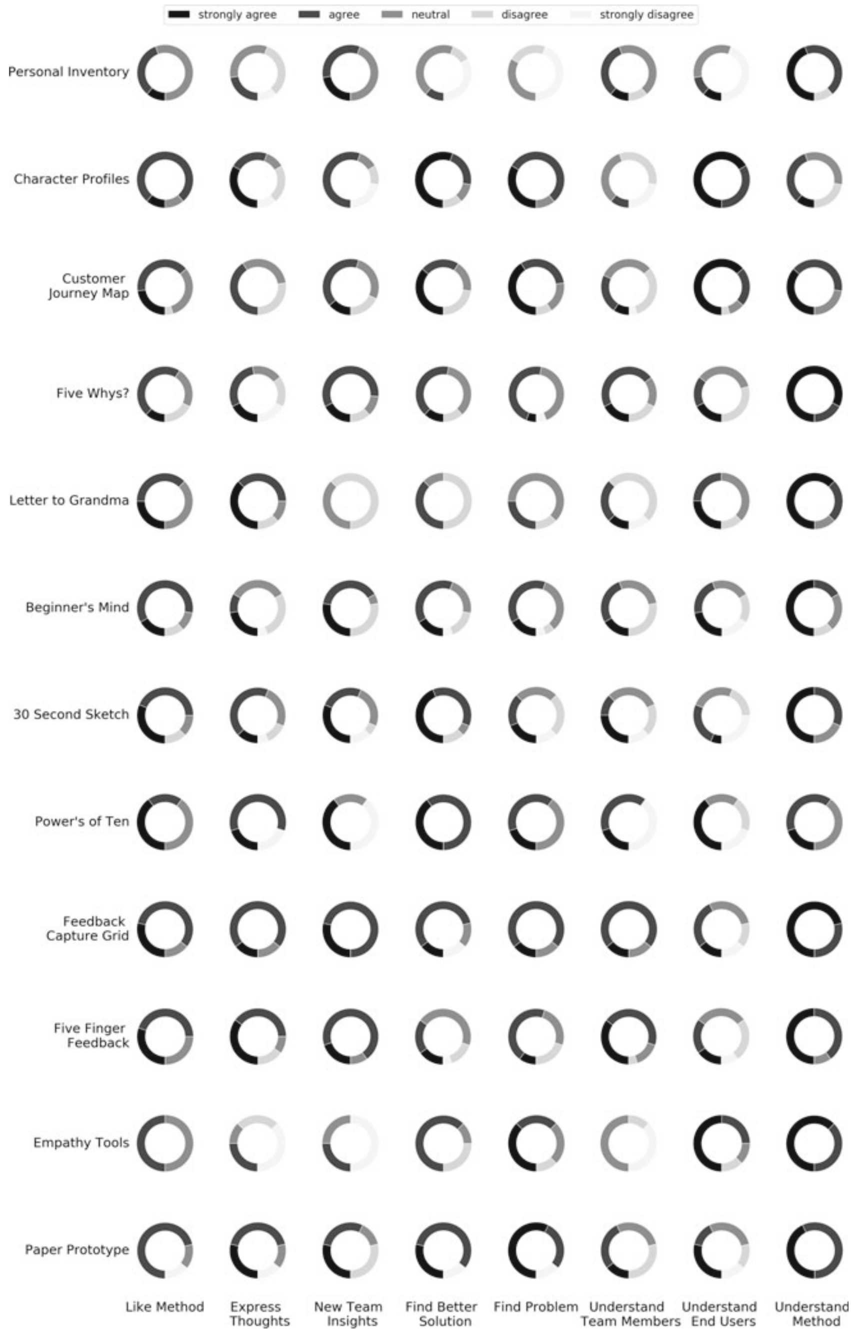


Fig. 2 Overview of method evaluation questionnaire results

Table 6 Methods with their benefits according to the participants

Method	Solving and finding problems	Communication	End user empathy	Team empathy
Personal inventory			X	X
Character profiles	X		X	
Customer journey map	X		X	
Five whys		X		X
Letter to grandma		X		
A beginner's mind	X	X		
30 Second sketch	X	X		
Powers of ten	X			
Feedback capture grid	X	X		X
Five finger feedback		X		X
Empathy tools			X	
Paper prototype	X	X	X	

Table 7 Methods with their possible applications as reported by the participants

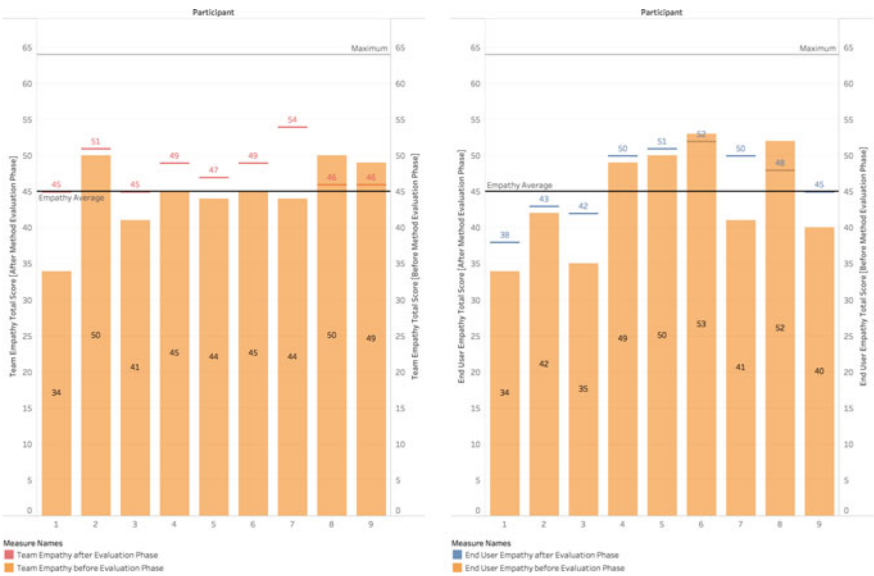
Method	Possible applications as mentioned by participants
Personal inventory	On-boarding to a project, team or employee
Character profiles	Find pain points in a system/verify and get ideas for new functionalities
Customer journey map	Find pain points in a system or solution and resolve these issues to improve UX
Five whys	When stuck/gathering customer requirements
Letter to grandma	Communicating new features and projects
A beginner's mind	When stuck/discuss new functionalities
30 Second sketch	UI design/ideation
Powers of ten	Understand a problem
Feedback capture grid	Gather feedback in any situation/test prototypes and solutions
Five finger feedback	Gather feedback in any situation/spontaneous one-on-one feedback/retrospectives
Empathy tools	Find UX problems in big systems, or systems targeting impaired people
Paper prototype	Designing new projects or features

Table 8 Mean empathy scores compared

	Mean team empathy	Standard deviation team empathy	Mean end user empathy	Standard deviation end user empathy
Initial assessment	44.67	5.05	44.00	7.21
Final assessment	48.00	3.04	46.56	4.80

5.2 *Empathy Evaluation*

We measured empathy using the adjusted Toronto Empathy Questionnaire for team and end user empathy, as described in Sect. 3. The total score has been calculated as defined by Spreng et al. (2009) and then compared for each participant. The mean of the total empathy scores were slightly below average (45) before the method evaluation phase, and rose to above average after application of the methods, as seen in Table 8. The average of 45 was defined by the original authors as the average empathy (Spreng et al. 2009). We use this definition to distinguish between lower than average empathy, and higher than average. The maximum possible total score is 64, (Spreng et al. 2009). We used the Wilcoxon Signed Ranks Test to calculate the significance of our results, in the same way as described by Patterson et al. (2017).



Team empathy End user empathy

Fig. 3 **a** Team empathy and **b** end user empathy before and after method evaluation phase

5.2.1 Team Empathy

As seen in Fig. 3a, the empathy towards other team members rose for most participants. Only in two cases there was a drop from 50 and from 49 down to 46 points. It should be pointed out, that all empathy scores are at least average after the method evaluation phase, where as before four participants had a lower than average empathy towards their team members. The average score is defined by the original Empathy questionnaire (Spreng et al. 2009).

There were two extreme boosts of eleven (17.18%) and ten (15.63%) points for participant one and seven, both attendees had an empathy score below average previously. Four moderate increases of four (6.25%) and three (4.96%) points were achieved by participants three, four, five and six. They all had an average or slightly below average score before the evaluation phase. Participant two gained one (1.56%) point on the scale, but this contributor already had an empathy score above average. Two participants had a moderate decrease of four (6.25%) and three (4.96%) points, nevertheless the score is still above average.

The Wilcoxon Signed Ranks Test shows that the increase is significant at $p \leq 0.05$. This is because the W-value is 8 and the critical value of W for $N = 9$ at $p \leq 0.05$ is 8.

Table 9 shows our coding categories, with subcategories and the mentions of the statements in the interviews. As can be seen, participants reported that team empathy has improved in all cases, five out of eight interviewees said that it improved significantly and could provide examples. It has been reported that employees learned how other roles work - or at least are more conscious of the differences now. More importantly, half of the participants declared that there is now more cooperation between roles, or at least they plan to involve other roles more into their daily work. More than half of the attendees claim to hold back fewer comments now, and one reports that he or she stopped holding back comments completely. Secondly, the collaborators reported that they are more talkative and conscious of their remote coworkers. In general, half of the interviewees declared that their team members know better how to communicate and two participants started talking with more people in their company.

5.2.2 End User Empathy

On the other hand, the empathy towards end users has risen in all but two cases, as seen in Fig. 3b. Participants three and seven had a huge rise of seven (10.94%) and nine (14.06%) points both originating from below average. Further, there were two moderate increases of 4 (6.25%) and 5 (7.81%) points for attendees one and nine from below average as well. Second, participants two, three, four and five had insignificant increases (1.56%) or decreases (−1.56%) of one point - one originated from below average, the others from above average. Finally, there was one moderate decline of four (6.25%) points for participant eight, from above average. Interestingly, attendee eight also had a decreased team empathy of four points, compared to that of before the evaluation phase. A possible explanation for that could be the effect of external factors, such as mood, health, etc. The Wilcoxon Signed Ranks Test shows the increase is significant at $p \leq 0.05$. This is because the W-value is 8 and the critical value of W for $N = 9$ at $p \leq 0.05$ is 8.

The findings from our solo interviews about end user empathy can be found in Table 10. Almost all participants reported that they now have a clearer picture of the end user and five could name an example. Likewise, half of the interviewees reported understanding requirements at least a bit better than before. One participant has now realised that there are requirements that do not fulfil the need of the end user and two further participants are now more aware of this problem. Two attendees claim to have started thinking in a human-centred way and another participant realises that problems in understanding customer requirements were a general issue and not his or her fault.

Table 9 Coding categories and mentions for team empathy from the interviews

Category	Answer	Mentions
Empathy	Absolutely improved, with examples	5
	Improved slightly, no real examples	3
Difference between roles	Learned how other roles work	5
	More conscious	2
	No difference, no change	1
Cooperation between roles	No change	4
	Much more cooperation	2
	More cooperation between roles	1
	More cooperation planned	1
Holding back comments	Holding back less now	4
	Still try not to hold back, no difference	3
	Held back, now says everything	1
General communication	People know better how to communicate with examples	4
	Started talking with more people	2
	No improvements	1
	Slightly, no examples	1
Remote communication	No improvements	4
	More talkative, more conscious	4

Table 10 Coding categories and mentions for end user empathy from the interviews

Category	Answer	Mentions
Clear picture of end user	Has increased in general, with example	5
	No change	1
	Improved for projects where methods were applied	1
	Has increased in general, without example	1
Understand all requirements	A bit better, no example	4
	No change	2
	More conscious	1
	Yes, if not asking	1
Requirements don't fulfil need	Yes, but no example	4
	Now more aware, asking more	2
	No	1
	Realized that this is happening, started to communicate with customer	1
Changed thinking	No change	4
	Started thinking in a human-centred way	2
	Changed thinking, no examples	1
	Realized that problems understanding requirements were a general issue, not his/her fault	1

6 Discussion and Limitations

Our study was based on the suggestion of several researchers that design thinking does not only provide value as a pre-development activity. As described in Sect. 2, some integrated processes propose using design thinking throughout the development process, either as a regular activity (Gurusamy et al. 2016; Glomann 2018; Sohaib 2018) or in an ad-hoc fashion (Dobrigkeit et al. 2018; Ximenes et al. 2015). Either way our DT@IT Toolbox provides support to development teams in such processes by suggesting and explaining suitable methods.

In our study the agile software team, developers and non-developers, were generally open towards using design thinking methods. This is in line with other research, e.g. Pedersen (2016). However, drawing or sketching seemed to be an uncomfortable activity to some participants (for example during the methods 30 Second Sketch and Paper Prototype). Possible solutions to this problem include, running warm-up exercises or providing a sketch cheat sheet, in order to offer easy ways to sketch user interfaces.

Even though our team was inexperienced in design thinking, they successfully applied the methods without further instruction and saw value and benefit in our toolbox. Similar to Pedersen (2016), who report that software developers were able to perform their work without the need of UX specialists, our team did not need the support of a design thinking specialist or coach. However, two of our participants were designers. Even though we did not perform any additional training, as was done by Pedersen (2016), and just handed out method sheets to the participants, our results suggest that this was enough to provide the attendees with a good understanding. Unlike Pedersen we did not modify most of our design thinking methods for the IT context but opted to provide IT related examples in our worksheets. This seems to have sufficed for our participants. However, as participants suggested our worksheets would benefit from more detailed and IT-related examples. For example, Character Profiles - the least understandable method in our study - had only one unspecific example. A detailed example with an exemplary persona and a task to be solved could help to get a better understanding of how to use Character Profiles. Likewise, it might be helpful to change the example of the Customer Journey Map to address an IT feature or project.

Concerning the suggested benefits of design thinking to agile software development our participants agree with current research and report of better communication within the team (cmp. De Paula et al. 2018; Clark and Smith 2008; Holloway 2009) increased empathy and understanding towards users and their needs (cmp. Brown 2008; Porcini 2009; Ward et al. 2009; Clark and Smith 2008) and improved problem and solution finding (cmp. Brown 2008; Martin 2010; Mahmoud-Jouini et al. 2016).

Even though most methods were perceived as useful and the participants had no problems understanding them, the pickup rate was rather low after the method evaluation phase. In our study we found two main reasons, a lack of time and a problem in realising that one of the methods would be helpful in a particular moment. The feeling of not having enough time for “unnecessary” activities has been reported before. For example, Gurusamy et al. (2016) found that people in IT development are concerned that too much time is spent before the development starts and Frye and Inge (2013), especially mention time spent on in feasible ideas. However, Frye and Inge (2013) also report that DT can eliminate much more waste than it creates if it is used correctly. While we can not directly show this from our results, we observed that various methods helped finding issues in solutions that were not developed yet. Thus, they reduced development effort by avoiding expensive redesigns after the implementation of a feature. To address the feeling of not having enough time, it could help to clearly state the value of specific design thinking methods and timebox them.

In order to address the issue of not recognising when a design thinking method could be helpful, knowledge about the available methods should be more prominent. This could be achieved, for example by hanging up method posters or regularly scheduling design thinking sessions to practice the methods. Additionally, we believe that design thinking methods can be integrated into agile meetings, e.g.:

- Backlog Grooming - 30 Second Sketch, Paper Prototype, Character Profiles, A Beginner's Mind, Powers of Ten
- Sprint Planning - Customer Journey Map, Letter to Grandma, Character Profiles
- Sprint Review - Feedback Methods, Empathy Tools
- Retrospective - Feedback Methods, Five Whys
- Interdepartmental Meetings - Personal Inventory

In terms of investigating whether regularly applying design thinking methods increases the participants' empathy towards the user and towards team members, our results show a significant increase. However, the increase in team empathy was noticeably higher than the increase in end user empathy. From our understanding, that happened for two reasons. First, most of the methods are designed to be used by inexperienced teams who want to increase communication and empathy. Second, applying the methods required team members to work together in collaboration, which additionally contributed to enhancing team empathy. Interestingly one of the participants showed a decrease in empathy, both towards the user and towards the team members. A possible explanation for that might be that external factors influenced the participant's results. However, since investigating the influence of other factors was not in the scope of this article, we could not verify what led participant eight to have a decreased level of empathy. Finally, another interesting fact, is that our results show that user empathy and team empathy do not always correlate among participants. For instance, participant two has a rather high score for team empathy, but end user empathy is below average.

7 Conclusion and Future Work

The goal of this study is to support agile development teams with beneficial design thinking activities. To that end, we created and evaluated the DT@IT Toolbox. We selected 12 methods from existing collections targeted at design thinking novices and created by established companies in the area of design thinking implementation. For our toolbox, we created a worksheet that explains the purpose of the method, how to implement it and provides examples that address software development activities for each of the 12 selected methods. We evaluated the toolbox with a team from DESIGN-IT, a medium sized software development company, over a period of 12 weeks. During that time, we analysed the benefits of the methods application and measured how the application period affected the team's level of empathy towards the user and within the team after all methods were applied. Our participants found most of the methods easy to understand and could apply them after reading our worksheets without the necessity of further training or coaches. Additionally participants saw value in the selected methods and could describe benefits as well as additional future use cases for all methods included in the DT@IT Toolbox. However, we noticed that only some participants used the methods after the study was conducted. After the evaluation period, empathy towards users and the team significantly increased for most participants. Our results are promising and suggest that DT methods add value to daily agile development and that a regular application can even increase team empathy and user empathy. However, we could only test a limited set of methods. For future work, we suggest analysing further methods. Specifically, the DT@IT Toolbox should be extended with observation methods, e.g. Shadowing (IDEO 2003) or Fly on the Wall (IDEO 2003), as participants reported they would like to see the software that they develop being used by the end users. Furthermore, there are several design thinking methods that could be used for a specific activity, e.g. various brainstorming or feedback techniques. For further research it would be interesting to test such methods against each other to evaluate which work best for agile teams. We also noticed that after the study, team members did not apply the methods as much as we expected. Accordingly, further research needs to be conducted in order to understand how to keep teams interested in using design thinking methods. A possible line of research could be to investigate whether team members are more willing to use shorter methods or whether tying methods to specific agile activities promotes application of the methods, e.g. make use of this method during reviews.

Our work extends existing research on integrating DT with agile software development methodologies by developing and validating a toolbox that software teams can use in their everyday work.

References

Aerssen, B., Buchholz, C., Burkhardt, N., Ernst, A., Rings, J., et al. (2018). *Das große Handbuch innovation: 555 Methoden und Instrumente für mehr Kreativität und Innovation im Unternehmen*. Vahlen.

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., et al. (2001). Manifesto for agile software development.

Belshee, A. (2005). Promiscuous pairing and beginner's mind: Embrace inexperience (agile programming). *Agile Development Conference (ADC'05)* (pp. 125–131).

Brauneck, P., Urbanek, R., & Zimmermann, F. (1995). *Methodensammlung: Anregungen und Beispiele für die Moderation*. Verlag für Schule und Weiterbildung: Dr.-Verlag Kettler.

Brown, T. (2008). Design thinking. *Harvard Business Review*, *86*(6), 84.

Carlgren, L., Elmquist, M., & Rauth, I. (2014). Exploring the use of design thinking in large organizations: Towards a research agenda. *Swedish Design Research Journal*, *1*(14), 47–56.

Clark, K., & Smith, R. (2008). Unleashing the power of design thinking. *Design Management Review*, *19*(3), 8–15.

Creswell, J. W., & Plano Clark V. L. (2017). *Designing and conducting mixed methods research*. Sage Publications.

Cross, N. (2003). *Engineering design methods: Strategies for product design* (Vol. 4).

d.School Stanford. (2018). Design thinking bootleg.

De Langhe, B., Bart, P., & Larrick, R. (2017). Linear thinking in a nonlinear world. *Harvard Business Review*, *95*(3), 130–139.

De Paula, D., Dobrigkeit, F., & Cormican, K. (2018). Design thinking capability model (DTCM): A framework to map out design thinking capacity in business organisations (pp. 557–566).

Dobrigkeit , F., & de Paula, D. (2019). Manifestations of design thinking within software development. In *Proceedings of ECIS19*.

Dobrigkeit, F., de Paula, D., & Uflacker, M. (2018). InnoDev—A software development methodology integrating design thinking, scrum and lean startup. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking—Research looking further: Design thinking beyond solution-fixation*, Understanding Innovation, (pp. 199–228). Springer.

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology*, *50*(9), 833–859.

Eames, C., & Eames, R.(1977). Powers of ten. (film). *Santa Monica: Eames Office*.

Erickson, J., Lyytinen, K., & Siau, K. (2005). Agile modeling, agile software development, and extreme programming: The state of research. *Journal of Database Management*, *16*(4), 88.

Ferreira, J., Noble, J., & Biddle, R. (2007). Agile development iterations and UI design. In *AGILE 2007 (AGILE 2007)* (pp. 50–58). IEEE.

Flick, U. (2014) *An introduction to qualitative research*, 5th ed. Sage Publications Limited.

Frye, U., & Inge, T. (2013) *The integration of design thinking and lean software development from the perspective of product owners and scrum masters*. Master's thesis, CHALMERS UNIVERSITY OF TECHNOLOGY.

Glomann, L. (2018). Introducing 'human-centered agile workflow' (HCAW)—An agile conception and development process model. In T. Ahram & C. Falcão (Eds.), *Advances in usability and user experience* (Vol. 607, pp. 646–655). Springer International Publishing.

Greenberg, S., Carpendale, S., & Marquardt, N., & Buxton, B. (2011). *Sketching user experiences: The workbook*.

Griffith, E. (2014). Why startups fail, according to their founders. *Retrieved September, 20*, 2016.

Grossman-Kahn, B., & Rosensweig, R. (2012). Skip the silver bullet: Driving innovation through small bets and diverse practices. *Leading Through Design*, 815–829.

Grudin, J., & Pruitt, J. (2002). Personas, participatory design and product development: An infrastructure for engagement. In *Proceedings of PDC* (Vol. 2002).

Gurusamy, K., Srinivasaraghavan, N., & Adikari, S. (2016). An integrated framework for design thinking and agile methods for digital transformation. In A. Marcus (Ed.), *Design, user experience, and usability: Design thinking and methods* (Vol. 9746, pp. 34–42). Springer International Publishing.

Hanington, B., & Martin, B. (2012). Universal methods of design: 100 ways to research complex problems. *Develop innovative ideas, and design effective solutions: Rockport Publishers*.

Hicks, A. (2015). *Empathy and experiential learning: How a community-based project enhances the college experience*. Ph. D. thesis, University Honors College, Middle Tennessee State University.

Hildenbrand, T., & Meyer, J. (2012). Intertwining lean and design thinking: Software product development from empathy to shipment. In A. Maedche, A. Botzenhardt, & L. Neer, (Eds.), *Software for People: Fundamentals, Trends and Best Practices*, Management for Professionals (pp. 217–237). Springer. https://doi.org/10.1007/978-3-642-31371-4_13.

Hitchcock, D., & Taylor, A. (2003). Simulation for inclusion... true user-centred design? In *Include 2003 Conference Proceedings* (pp. 105–110).

Holloway, M. (2009). How tangible is your strategy? How design thinking can turn your strategy into reality. *Journal of Business Strategy*, *30*(2/3), 50–56.

IDEO. (2003). *IDEO method cards: 51 ways to inspire design*.

Jalleh, S. (2013). Co-designing an EC135 air ambulance cabin, p. 8.

Jobst, B., & Meinel, C. (2013). How to give orientation and inspiration to design thinking novices regarding prototyping methods? (p. 10).

Kollmann, J., Sharp, H., & Blandford, A. (2009). The importance of identity and vision to user experience designers on agile projects. In *2009 Agile Conference* (pp. 11–18). IEEE.

Kumar, V. (2012). *101 design methods: A structured approach for driving innovation in your organization*. New York: Wiley.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing, 80*(6), 69–96.

Leung, W. -C. (2001). How to design a questionnaire. *student BMJ, 9*.

Lewrick, M., Link, P., Leifer, L., & Langensand, N. (Eds.). (2017). *Das design thinking playbook: mit traditionellen, aktuellen und zukünftigen Erfolgsfaktoren*. Vahlen. OCLC: 958007004.
Lindberg, T., Meinel, C., & Wagner, R. (2011). Design thinking: A fruitful concept for IT development? In C. Meinel, L. Leifer, & H. Plattner (Eds.), *Design Thinking, Understanding Innovation* (pp. 3–18). Springer. https://doi.org/10.1007/978-3-642-13757-0_1.
Mahmoud-Jouini, S. B., Midler, C., & Silberzahn, P. (2016). Contributions of design thinking to project management in an innovation context: Contributions of design thinking to project management in an innovation context. *Project Management Journal, 47*(2), 144–156.
Martin, R. (2010). Design thinking: Achieving insights via the "knowledge funnel". *Strategy & Leadership*, *38*(2), 37–41.
Meinel, C., Rhinow, H., & Köppen, E. (2013). Design thinking prototyping cardset. *Hasso Plattner Institut für Softwaretechnik*.
Miaskiewicz, T., & Kozar, K. A. (2011). Personas and user-centered design: How can personas benefit product design processes? *Design Studies, 32*(5), 417–430.
Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing Research, 40*(2), 120–123.
Moser, C. (2012). *User experience design: Mit erlebniszentrierter Softwareentwicklung zu Produkten, die begeistern*. Berlin Heidelberg: X.media.press. Springer.
Ohno, T. (1988). *Toyota production system: Beyond large-scale production*. CRC Press.
Patterson, L. E., Boone, A. E, & Engsberg, J. R. (2017). The creation and testing of a simulation-based workshop to increase level of understanding and empathy of eating disorders. *Journal of Psychiatry and Psychiatric Disorders, 1*(5), 270–289.
Pedersen, T. Ø. (2016). *UX Toolbox for Software Developers*. Ph.D. thesis.
Pérez, J., & Murray, M. C. (2010). Generativity: The new frontier for information and communication technology literacy. *Interdisciplinary Journal of Information, Knowledge, and Management*, *5*, 127–137.
Pojasek, R. B. (2000). Asking? Why?? Five times. *Environmental Quality Management*, *10*(1), 79–84.
Porcini, M. (2009). Your new design process is not enough-hire design thinkers!. *Design Management Review*, *20*(3), 6–18.
Ritchie, J., Tinker, A., & Power, J. (2016). The effectiveness of design thinking techniques to enhance undergraduate student learning. In *Futurescan 3: Intersecting identities*. Association of Fashion & Textiles (FTC), Loughborough University.
Rosenbaum, M. S., Losada Otalora, M., & Ramírez, G. C. (2017). How to create a realistic customer journey map. *Business Horizons, 60*(1), 143–150.
Schwaber, K., & Beedle, M. (2001). *Agile software development with scrum*.
Serrat, O. (2017). *The five whys technique* (pp. 307–310). Springer Singapore.
Snyder, C. (2003). *Paper prototyping: The fast and easy way to design and refine user interfaces*. Morgan Kaufmann.
Sohaib, O., Solanki, H., Dhaliwa, N., Hussain, W., & Asif, M. (2018). Integrating design thinking into extreme programming. *Journal of Ambient Intelligence and Humanized Computing*.
Sonnenwald, D. H. (1995). Contested collaboration: A descriptive model of intergroup communication in information system design. *Information Processing & Management, 31*(6), 859–877.
Spreng, N. R., McKinnon, M. C., Mar, R. A., & Levine, B. (2009) The Toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment, 91*(1), 62–71.
Suzuki, S. (1970). *Zen mind, beginner's mind*.
Tam, M. (2001). *Using paper prototyping as a usability testing methodology for web application development*. Master's thesis, School of Interactive Arts and Technologies—University of British Columbia.
Tashakkori, A., & Teddlie, C. (2010). *Mixed methodology: Combining qualitative and quantitative approaches*. Sage Publications Inc.

Valencia, A., Person, O., & Snelders, D. (2013). An in-depth case study on the role of industrial design in a business-to-business company. *Journal of Engineering and Technology Management*, *30*(4), 363–383.

Verganti, R. (1997). Leveraging on systemic learning to manage the early phases of product innovation projects. *R&D Management*, *27*(4), 377–392.

Ward, A., Runcie, E., & Morris, L. (2009). Embedding innovation: Design thinking for small enterprises. *Journal of Business Strategy*, *30*(2/3), 78–84.

Ximenes, B. H., Alves, I. N., & Araújo, C. C. (2015). Software project management combining agile, lean startup and design thinking. In A. Marcus (Ed.), *Design, user experience, and usability: Design discourse* (Vol. 9186, pp. 356–367). Springer International Publishing.

Poirot: A Web Inspector for Designers

Kesler Tanner, Naomi Johnson and James A. Landay

Abstract To better understand the issues designers face as they interact with developers and use developer tools to create websites, we conducted a formative investigation consisting of interviews, a survey, and an analysis of professional design documents. Based on insights gained from these efforts, we developed Poirot, a web inspection tool for designers that enables them to make style edits to websites using a familiar graphical interface. We compared Poirot to Chrome DevTools in a lab study with 16 design professionals. We observed common problems designers experience when using Chrome DevTools and found that when using Poirot, designers were more successful in accomplishing typical design tasks (97–63%). In addition, we found that Poirot had a significantly lower perceived cognitive load and was overwhelmingly preferred by the designers in our study.

1 Introduction

More designers are finding themselves with a need to delve into the role of developer as end-user programmers, whether due to a lack of developer bandwidth in an organization, a desire to communicate visual details to a developer, or a need to optimize a design due to the difference in rendering between the original design tool and a web browser (Rosson et al. 2005; Dorn and Guzdial 2006). For many designers, this foray from their world of visual design into a world of code and syntax is mandatory, but neither instinctive nor necessarily welcome (Dorn and Guzdial 2010a).

Amiri has described the dissonance experienced by many designers working to understand development. He compares programmers to linguists, trained profes-

K. Tanner (✉) · N. Johnson · J. A. Landay
Department Computer Science, Stanford School of Engineering,
475 Via Ortega, Stanford, CA 94305, USA
e-mail: keslert@gmail.com

N. Johnson
e-mail: snj3k@virginia.edu

J. A. Landay
e-mail: landay@stanford.edu

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_14

sionally to understand the "syntax, semantics and pragmatics" of a language, while designers, in contrast, are more akin to tourists, interested in the language of coding only in so far as it is needed to "communicate with people[,] …explore the new environment[,] and …get by" (Amiri 2011).

There are tools to aid designers in navigating the developer space. Some of the better-established tools, however, denote by their very names that they have been lent from the developer's toolbox and were not created with the designer in mind. These are native browser web developer tools, such as Chrome DevTools, Firefox Firebug, and Safari Inspector.

Alongside these native web development tools, researchers have made new developer tools targeted at end-user programmers. WebCrystal (Chang and Myers 2012a) and CopyStyler (Fitzgerald et al. 2008) are both tools developed to assist novices. However, these tools focus on educating the novice in the comprehension and use of syntax, rather than on leveraging a designer's unique strengths.

Additional tools extend the capabilities of native web tools. While these systems could potentially be helpful to designers, they often address issues far beyond the capabilities of many designers. For example, while these tools are tackling issues like animation and interactivity (Hibschman and Zhang 2016a; Burg et al. 2015a), designers are stumbling on much simpler problems like changing colors and font sizes. These problems might be trivial for a developer but are nonetheless hindrances to a designer.

In limited areas, the disconnect between development and design has been alleviated through web design tools such as Webflow, Adobe Dreamweaver, and other WYSIWYG editors. These tools allow a user to create a website using interfaces that are closer to traditional graphic design tools (Rode et al. 2005). However, while these tools work well in small productions, such as a personal blog, they are not solutions typically used in professional environments, where developers prioritize maintaining greater control of their website code (Rosson et al. 2005).

Therefore, new tools are needed that cater to the designer. For these new tools to be effective, a better understanding of the interaction between designer and developer in the modern professional ecosystem is critical. To help address these problems, we conducted interviews with several professional designers and developers to understand how they worked together to solve web development tasks. Based on the insights from these interviews, we created a survey and garnered responses from 43 professional UI/UX designers. Finally, we collected and analyzed professional design documents to confirm the findings of our survey.

Building on related work and our findings from our formative investigation, we designed and implemented Poirot (Fig. 1), a web inspector tailored for designers that enables them to make style edits to websites using a familiar graphical interface. Poirot enables users to select an element on a website through direct manipulation, which displays a panel showing the element's current styles and allows modification of those styles. The tool assists the user in making consistent style choices by constraining the available options to those existing in the website's design system. Poirot remembers all updates to the website and can toggle back and forth between the original version and the modified version. When the user is satisfied with the

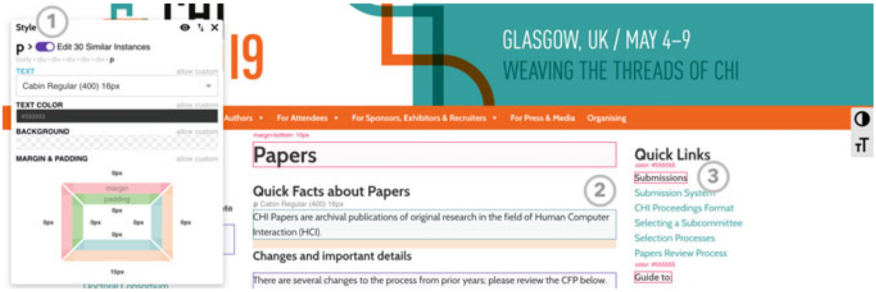


Fig. 1 (1) The Poirot web inspection panel. (2) The element selected is highlighted in blue with a label providing additional details. Similar instances are highlighted in purple. (3) Previous changes are highlighted in red with labels showing changed properties and values

results, Poirot can export these changes as a JSON file that can be imported by a developer for a live preview of the changes.

We evaluated the effectiveness of Poirot in a lab study with 16 professional UI/UX designers. We discovered that with no training, designers were able to successfully complete more tasks using Poirot than with Chrome DevTools (DevTools).

The contributions described in this paper are:

- A formative investigation of designers' web development experience that show that designers have close and frequent interaction with developers, often modify existing web interfaces, and play a leading role in maintaining the visual quality of websites.
- Poirot, a novel web inspection tool for updating the styles of existing websites, allowing designers to modify websites using a familiar graphical interface.
- A within-subjects evaluation of Poirot compared to a popular web inspection tool used for live edits of websites. This evaluation shows that when using Poirot users had significantly higher task completion rates, faster task completion times, a lower perceived cognitive load, and an expressed preference for Poirot interface to Chrome DevTools.

2 Related Work

Prior research reveals that nonprogrammers approach development differently than programmers (Rode and Rosson 2003). Indeed, web development, a growing form of end-user programming, has presented many challenges for teachers and curriculum makers (Wang and Zahadat 2009; Ko et al. 2004; Rosson et al. 2005; Chilana et al. 2015; Blackwell 2002). Dorn studied graphic designers as end-user programmers, looking to address these difficulties (Dorn and Guzdial 2006). His research provides insight into graphic designers' knowledge, understanding, and perspectives of programming (Dorn and Guzdial 2010a, b). Park builds on this research, offering additional insight into the difficulties and errors novices experience when learning HTML and CSS (Park et al. 2013b, 2015). This collective research provides a foundation for understanding end-user programmers and has led to the creation of novel tools.

For example, OpenHTML is an interface that improves a novice's experience when writing HTML and CSS (Park et al. 2013a). FireCrystal (Oney and Myers 2009), Theseus (Lieber et al. 2014), Clematis (Alimadadi et al. 2014), and Telescope (Hibschman and Zhang 2016b) help users identify code causing specific output. Scry (Burg et al. 2015b) and Unravel (Hibschman and Zhang 2015) help users reverse-engineer the cause of behaviors in their code. Ply reduces complexity by hiding irrelevant code (Lim 2017). Tutorons provides inline explanations of CSS and HTML code (Head and Hearst). WebCrystal (Chang and Myers 2012b), Copystyler (Fitzgerald et al. 2008), C3W (Fujima et al. 2004), and Marmite (Hong and Wong 2006) enable users to combine code from different websites to assist novices to create mashups of ideas. Chickenfoot allows users to insert, remove, and replace functionality of live websites without access to the source code (Bolin et al. 2005).

While these web development tools can be helpful to designers, they often require the user to be actively involved in the code. Designers have varying degrees of technical experience, and while some would identify as novice developers, not all designers are interested in using tools that were designed to improve the users' coding ability. Like existing tools, Poirot seeks to help designers make changes to websites. However, rather than encouraging or educating designers to learn more about development, it works to users' strengths by allowing them to modify websites through a graphical interface.

3 Formative Investigation

As Goodman et al. noted, there is a "mismatch between HCI research and design practices" (Goodman et al. 2011); this has been the case for decades (Gould and Lewis 1985; Putnam et al. 2016a) and has been discussed in various papers about interaction design (Putnam et al. 2016b; Goodman et al. 2011; Wolf et al. 2006). To better understand how designers and developers cooperate in accomplishing their web development tasks, we interviewed four professional designers and developers working for multiple technology companies. Based on our insights from those interviews, we surveyed 43 UI/UX professionals working for a large technology corporation to better understand their design activities. Finally, we collected design documents and analyzed the types of changes requested by designers in those documents. These studies increased our understanding of tasks designers struggle to complete and highlighted new tool functionality that would be helpful to designers.

3.1 Interviews

We interviewed four professional designers and developers (two female, two male). Two identified as UI/UX designers, one identified as a software engineer, and the fourth was transitioning from the role of UI/UX designer to software engineer. All worked at large technology corporations in the United States. These interviews were unstructured and lasted up to an hour. One was conducted in-person while the others were done over Skype. Participants were not compensated for their time. During the interviews, we asked open-ended questions regarding the participant's current role. We also asked designers about situations in which they work with developers, and we asked developers about situations in which they work with designers.

From our interviews, we gained several insights into the designer/developer ecosystem in professional settings. First, we learned that designers and developers interact frequently, often on a daily basis, to accomplish design related tasks. Designers rely on developers to implement their designs and solve related problems. However, designers will choose to struggle through using inspection tools (e.g., DevTools) to solve issues on their own before interrupting a developer. If a designer is unable to solve the issue, she or he relies on help from a developer. An illustration of these close and frequent interactions is provided in the following example, which is a composite of what we heard in our interviews.

A designer starts creating a design mockup in a graphics tool. In these early stages, the designer converses with her developers to ensure her designs are feasible. When the designer completes the mockup, she adds additional markings (usually in red) to the design, which specify details about different design elements, like the number of pixels between items or the size and color of a heading. The designer then hands this "redline" document over to a developer, who transforms it into code for the website. During this phase, the developer and designer have several additional conversations to clear up any questions and add additional details.

After the design is implemented, it is published to the live site or staging server. At this point, the designer previews the design in her browser and confirms that the implementation was satisfactory. If there are issues, the designer attempts to fix them using DevTools. In situations where she is able to fix the problems herself, she documents the changes and communicates them to the developer. Depending on the size and number of issues, this communication happens verbally, by email, or through formal logging in a bug tracker, and is often accompanied by screenshots and/or the revised CSS or HTML updates.

In cases where the designer is unable to make the corrections herself using inspector tools, she then tries to find a time when the developer can work side-by-side with her. This enables the designer to watch the developer navigating DevTools while she explains what changes she wants to see. If the designer cannot get time with a developer, she logs the issue in a bug tracker without a documented solution.

Through our interviews, we also observed the efforts of designers and developers to maintain a site's integrity. Developers normally take the lead on code quality of a website, while designers oversee the visual quality of a website. Issues discovered during perusal of the site are treated with the same fix-and-document pattern described above. For example, one participant described a recent large and "grueling" undertaking in which he performed an audit of his company's entire site looking for instances where the company's design system (Frost 2016) had not been followed. This audit consisted of browsing through the site looking for visible issues, as well as searching through the code base for the use of custom classes where these abuses often occurred.

3.2 Designer Survey

To confirm and deepen the findings from our interviews, we designed an online survey to gather feedback from a larger set of UI/UX designers. The survey was sent out on an internal Slack channel at a large technology company. Participants were entered into a raffle for a $50 Amazon gift card in exchange for their participation. Forty-three UI/UX professionals completed the survey. They had an average 6.6 (sd = 5.7, min = 2, max = 25) years of experience as professional designers. Their average level of design expertise as defined by "How would you define your level of visual design expertise?" on a 7-point Likert scale from Novice to Expert was 5.5 (sd = 1.1).

All participants reported having created UI design mockups as part of their job, and 36 said they regularly create UI design mockups. To the question, "What is your experience with creating redline design documents to give to a developer?", 21 reported that they regularly create redline documents to give developers, 15 reported that in the past they had created redline documents for developers, 5 reported that they were familiar with redline documents but had never created one, and 2 reported having never heard of redline documents.

On a 7-point Likert scale from "Strongly disagree" to "Strongly agree", participants averaged 6.2 (sd = 1.1) for how often their work includes making improvements to existing user interfaces. To understand what kind of changes designers typically made, they were asked to select all changes they might realistically make when modifying a user interface. Table 1 shows the selection percentages.

Finally participants were asked, "In what aspects of your work do you interact with developers? What form does this interaction take?" These free responses confirmed

Table 1 Edits that designers from our survey realistically make when modifying a user interface

Task	Selected (%)
Add/decrease whitespace between elements	98
Change the text size	98
Change the text font	93
Change the text color	88
Change the background color	88
Change the text copy	88
Replace an image with another image	88
Add or modify a drop shadow	81

our previous findings regarding developer/designer relations and interactions. The following participant's statement is reflective of answers provided:

> Basically daily. They work on the designs I produce. I show them design stuff. Ideally we have a conversation about it (though not always unfortunately.) They work on it and come to me with questions (or just implement it how they like and wait for the QA later). I do QA on it.

3.3 Collection and Analysis of Design Documents

To verify that the types of changes designers said they made in the survey were actually what we saw in practice we collected over 50 redline design documents from a series of Google Image searches for keywords like "redline design document." Many of these documents came from technical write ups on UI designers' portfolio sites. We analyzed these documents for changes requested and notations used, categorizing the types of changes we found. While the style of redlining differed greatly, there was a commonality of changes. These changes supported the answers we received from the survey and interviews and most often included edits of color, content, shadows, typography, and whitespace.

3.4 Key Insights

Our investigation provided us with the following insights. These insights influenced our tool design and should be helpful to other future design efforts: First, discrepancies arise during the translation between the designers' mockups and the developers' HTML/CSS implementation, and designers often drive the efforts to correct these discrepancies. Second, designers struggle to remember CSS syntax and language. Third, designers struggle to know what elements are going to be affected by a change when modifying CSS. Fourth, designers tediously document their changes in an effort to minimize translation errors, and this documentation often requires the use of foreign terminology. Fifth, significant efforts are made to keep a website and design system in sync.

4 Tool Design

In this section we describe how we used the insights from our formal investigation to design and implement Poirot (Fig. 1), a web inspector for designers.

In on our investigation, we found that errors frequently occurred during translation between the designers graphical mockup and the developers HTML/CSS. As a result, designers often attempted to use DevTools to correct these errors, but they struggled to effectively use this developer tool. These problems suggested that our tool should target the final medium but provide a user interface and interactions that were familiar to designers. For this reason, we built Poirot to work as a bridge between the designer and developer mediums, allowing designers to modify a live website using a familiar graphics interface while producing results that are already in the medium of the browser. This permits designers to make the same pixel-precise changes they can make in graphics tools in the browser without worrying about their work getting lost during the translation to code since the translation is occurring in realtime while they work.

Poirot functions as a Chrome extension that can be injected into a page by clicking on the Poirot icon in the extensions bar. When Poirot is first loaded into the page, it takes inventory of all of the HTML elements, assigns each a unique ID, and categorizes them into text elements, image elements, and regular elements. These classifications are used to adapt the user interface and selection algorithms.

Poirot overrides default styling through CSS selector precedence by injecting a new stylesheet into the page and writing highly specific selectors (#poirot#poirot#poirot .poirot-256) with each CSS declaration using the !important tag. Unless a website is using highly discouraged !important inline declarations, Poirot selectors will have precedence.

4.1 Selecting Elements

Our formal investigation showed that when making CSS edits, designers struggled to know what elements were going to be affected. Poirot allows for a familiar graphical element selection experience. As the user hovers over elements, a semi-transparent gray box outlines the hovered element and a small label appears showing the element's HTML tag name. For text elements, the label also includes font family, font weight, and font size (Fig. 1). When the user selects an element by clicking on it, the element's bounding box is highlighted in semi-transparent blue and the Poirot panel updates to show the properties of the element.

Poirot also shows other similar instances that will be edited *along* with the currently selected item. Poirot has different strategies for determining similar instances based on the type of element selected. If the element is a text element, Poirot finds similar instances throughout the page that have the same color, background color, font size, font family, font weight, and padding. Otherwise, Poirot looks for similar instances that are siblings of the selected element in the HTML structure and share background color, box shadow, margin, and padding, or elements throughout the page that have the same bounding box dimensions. These algorithms for determining similar instances worked well on the websites we tested, but they could be improved with even more targeted algorithms or machine learning.

These similar, auto-selected instances are outlined in a purple highlight. If the user desires to edit only the manually selected element, the auto-selection can be toggled off, and only the manually selected element will be edited.

4.2 Updating Styles and Content

Our formal investigation highlighted that designers struggle with remembering CSS syntax and language. Poirot reduces much of the complexity of updating styles and content by eliminating the need to remember CSS syntax and by handling logistical details for users.

Rather than each style being a property-value pair in a CSS rule, Poirot displays properties as UI widgets. As elements are selected, the Poirot panel updates the widgets it shows to display only those that are relevant. For example, if the element is an image element, the panel displays a "Change Image" button. If the element is a text element, it displays a typography dropdown selection (Fig. 1).

For the browser to properly display fonts, the font needs to be installed locally on the machine, or the page needs to fetch the font file. Poirot handles this complexity by allowing the user to select the desired font from a list, then automatically makes the request for the font file. Currently only fonts served by Google Fonts are supported, but it could be extended to include other font services.

The same is true for image exploration. To display an image, the image needs to have a valid URL, either by being uploaded to a server or encoded as a Data URI. In Poirot, users can swap out an image for another by clicking on an image element and selecting the "Change Image" button. This opens a native file picker where the user can select the desired image. When the image is selected, Poirot converts the image to a Data URI and substitutes it for the current image.

Using standard copy/paste hotkeys, Poirot allows users to copy styles from one element and paste them onto another element. When the styles are pasted, a window shows the different properties that are being transferred (Fig. 2). As toggles are turned on or off, the element receiving the style transfer updates in real time. When the user is satisfied with the transfer they can click to apply the styles. When a style transfer is complete, the user can choose a new element to act as a copy source and the same recipient element can be styled again. In this manner, the user could transfer the font size and font family from one element, the color from another element, and the margin and padding from a third element. Copying styles even works across elements from different websites, which is a useful ability since web designers often look at other people's websites, select the elements they like, and combine these in their own designs (Herring et al. 2009).

4.3 Maintaining Visual Consistency

Our formal investigation showed that significant effort is required to keep a website and design system in sync. To aid in this effort, Poirot has built-in support for design systems and currently supports colors, typography, shadows, and spacing (Frost 2016). Poirot's UI widgets help designers make consistent choices by only presenting

Fig. 2 Poirot allows copying styles from one element and pasting the styles to another element. The resulting window lets designers toggle on and off properties to see a live preview of what styles are being transferred

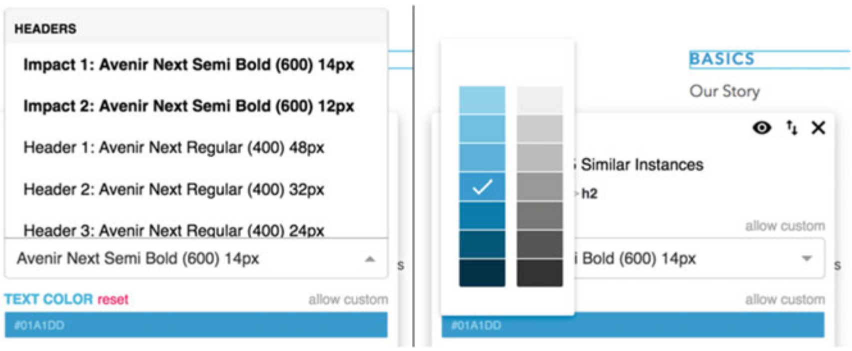


Fig. 3 UI widgets for typography and color showing the built in design system choices

values from the design system (Fig. 3). If desired, users can break out of the design system by toggling "Allow custom" next to a UI widget.

In addition to helping the user make consistent choices when manually updating the page, the built in design system supports an interactive design system audit of a page. Figure 4 shows an audit of the typography used on a sample page. By turning on "Header", all header text elements are highlighted in green. Instead of turning on all headers, a specific type of header can be turned on, such as "Header 1". Under each type is a list of the elements using that type. Clicking on an element navigates to the element on the page.

The audit also shows *unknown* typography not defined by the design system. By going through these elements, a designer can decide whether to update the design

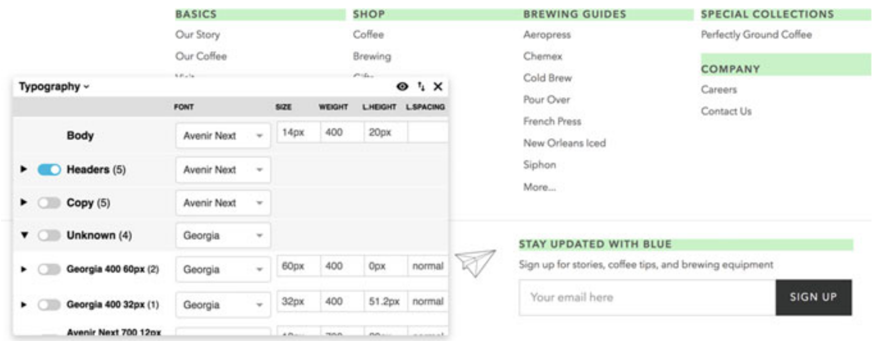


Fig. 4 The typography panel allows the designer to perform an audit to see where typography is being used and any unknown fonts that are not in the website's design system

system to include this unknown font or exchange the font for one defined in the design system.

4.4 Tracking Changes and Documentation

Our formal investigation revealed the tedious work of documenting requested website updates. Poirot eliminates this work by tracking all changes that are made to a page by a designer. When a property changes, the UI widget displays a "reset" button that allows the user to revert to the original value. Poirot also provides functionality to toggle on and off all modifications, allowing the user to quickly compare and contrast the updated page with the original page.

Page changes can be outlined by turning on "Redline changes". This highlights all changes in red and shows a label with the properties that changed and their new values (Fig. 1). This notation mimics the notation styles we saw in our document analysis. When a designer is satisfied with their changes, they can export them as a JSON file to share with a developer or attach to a bug report. The developer can import the JSON file, turn on redlines, and visualize the updates. As she makes changes in the code that modify these values to match those made by the designer, Poirot will no longer highlight these elements. In this manner, developers can use the redline functionality as a checklist for making updates. When there are no more redlines, the developer has implemented all of the designer's changes.

5 User Study

We evaluated Poirot in a within-subjects lab study in which participants used Poirot or DevTools to accomplish a series of design tasks. We selected DevTools as our comparator based on its popularity and feedback from our survey. For each task we measured task completion and task time.

5.1 Participants

Sixty-four participants volunteered for the study. All were recruited using convenience and snowball sampling from Facebook, Twitter, Slack, and Nextdoor. Recruitment messaging stated that we were seeking UI/UX designers. Participants filled out a pre-survey questionnaire containing questions about their design experience. Thirty-six participants said they had worked in the role of UI/UX designer, were comfortable or extremely comfortable using Sketch or Photoshop, and had created UI mockups or regularly created UI mockups. We emailed these 36 participants, and 16 participants signed up and participated in the in-person study. Nine were female. Participants reported an average 7.2 (sd = 4.6, min = 2, max = 18) years of experience as UI/UX designers. Four participants reported being extremely comfortable with DevTools, seven reported being comfortable, and seven had used DevTools but considered themselves beginners. Two participants reported being extremely comfortable with code editors, seven reported being comfortable with code editors, and seven reported using code editors but considered themselves beginners. All participants reported being familiar with design systems. The study lasted 60 min and participants received a $75 Amazon gift card.

5.2 Apparatus

The study was conducted on a MacBook Pro (Retina, 15-in., Late 2013) laptop using the Google Chrome browser Ver. 67. Fifteen participants used the built-in trackpad and one used an optional external mouse that was provided. Scroll direction was adjusted to "natural" or "unnatural" to match the participant's normal scroll usage, and DevTools was positioned to "dock to right," "dock to bottom" or "undock into separate window" based on participant preference.

5.3 Tasks

The study consisted of nine design tasks. These tasks were selected based on our formative investigations, that showed that making these types of changes to existing user interfaces was common among UI/UX designers. As further confirmation, we asked study participants in a post survey how representative the study tasks were of actual tasks they might attempt to do in DevTools. On a 5-point Likert scale of "Not very representative" to "Very representative", they gave an average rating of 3.9 (sd = 1.1). The tasks were as follows:

1. Change the text color from gray800 to blue500.
2. Change all four gray800 button backgrounds to blue500.
3. Add a shallow drop shadow to both images.
4. Change the text to "The Latest from Blue Bottle".
5. Change the image to shop.jpg (file located in ~/Dropbox/Blue/).
6. Change both kicker fonts from Impact 2 to Impact 1.
7. Change the quote font family from Georgia to custom font Chelsea Market.
8. Increase the vertical whitespace between the text and button from 10 to 40 px.
9. Decrease the vertical whitespace between links from 10 to 0 px.

These design changes were all made on the same website. The website used for the study was a coffee shop website that we felt was representative of a modern website. It had a height of 3766 pixels, and the browser window on the study laptop showed 803 pixels, resulting in 4.7 viewports.

To help participants understand where to make changes to complete each task, they were provided with a $22' \times 8.5'$ paper that showed a print version of the website as it should look after all the changes had been made. Next to each of the changes was a numbered red circle that indicated the corresponding numbered task.

In addition, participants were given a paper with a simple design system to use (Fig. 5).

Our hypothesis was that Tool (Poirot/DevTools) would have significant effect on task completion and task time. Further, we wanted to explore task strategy as a moderator. We realized that tasks could be grouped into two categories based on the strategy that could be used to solve them: Find and Replace (F&R) and Custom. If a task was classified as F&R, it meant that a CSS rule existed targeting the exact set of elements that needed to be updated and the CSS rule contained a CSS declaration with the property that needed to be updated. If the declaration did not exist or the CSS rule targeting the set of elements did not exist, the task was categorized as Custom. Six of the nine tasks (2, 4, 6, 7, 8 and 9) were F&R tasks.

5.4 Procedure

Participants received a document describing the study procedure and introducing the design tasks, the design system, and the printed version of the website.

Participants were given three minutes to complete each task with up to an additional 30 s if they were in the middle of an attempted solution. Participants could skip a task at any time. They were also told they could use Google or any other online resource to complete the tasks. No training was provided for how to use either tool. The order in which participants used the tools was counterbalanced.

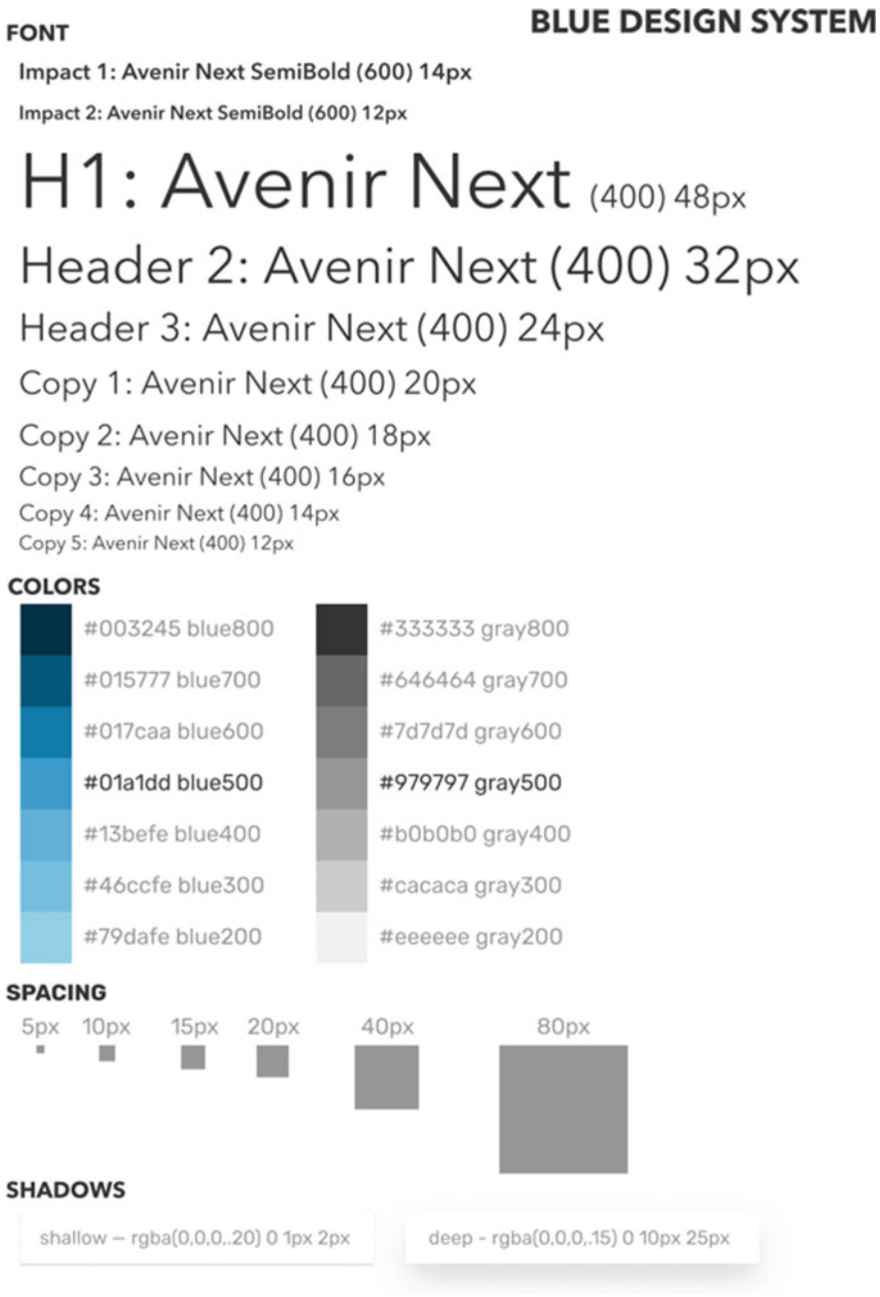


Fig. 5 Design system used by participants

When a participant believed they had completed a task, they removed their hands from the keys and either raised their hand or verbally said “Done.” The timer was then stopped and their work was checked. If it was correct, the time was recorded and the participant was asked to move to the next task. If it was incorrect, the participant was told what was incorrect, and the timer would continue. After attempting all nine tasks, the participant completed a standard NASA TLX questionnaire (excluding physical demand). Afterwards, the website was reset to its original state and the participant again attempted the nine tasks, this time with the alternate tool. After undertaking the nine tasks for the second time, the participant would again complete the NASA TLX questionnaire. Finally, participants completed a post-survey questionnaire asking about their experience with both tools. Following the survey, followup questions were asked to better understand the experience of using both tools, how participants used DevTools in their current work, and how they typically worked with developers.

6 Results

We present our results as a function of Tool (Poirot vs. Devtools) and Strategy (F&R vs. Custom). The values on each bar in Figs. 6 and 7 denote the number of data points.

6.1 Task Completion

To examine the effect of Tool on task completion, a logistic repeated-measures regression was conducted (Fig. 6). Logistic regression was used since the outcome variable is dichotomous. There was a significant effect of Tool (b = 4.11, $p < 0.01$) indicating that the odds of completing a task was 20.36 times larger when using Poirot.

A logistic repeated-measures regression was conducted to examine the effect of Strategy on task completion (Fig. 6). There was a significant effect of Strategy (b = -2.63, $p < 0.001$) indicating that the odds of completing a task was 6.61 times larger on F&R tasks.

Furthermore, a logistic repeated-measure regression was conducted with F&R tasks to examine the impact of Tool on task completion. There was no significant effect of Tool (b = 4.89, $p > 0.05$) indicating that for tasks that required F&R solutions, the odds of successfully completing a task was similar regardless of the tool used.

Another logistic repeated-measure regression was conducted with Custom tasks to examine the impact of Tool on task completion. The model failed to converge due to insufficient sample size, but the raw data provides suggestive evidence. The probability of completing a task was only 20.8% when participants used DevTools, compared to 97.9% when participants used Poirot.

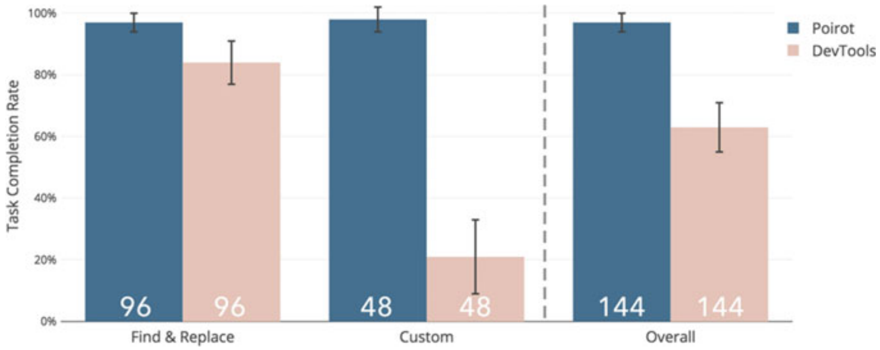


Fig. 6 Average task completion rate per solving strategy and overall. Error bars represent 95% confidence intervals

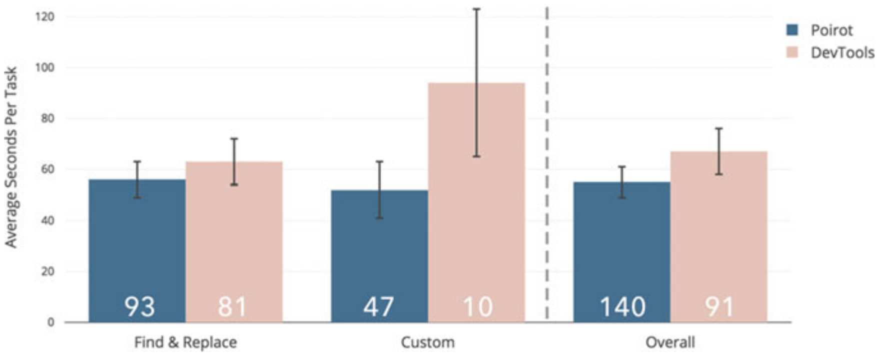


Fig. 7 Average task completion time per solving strategy and overall excluding failed tasks. Error bars represent 95% confidence intervals

6.2 Time

A Tool by Strategy repeated measures ANOVA was conducted on the time participants spent on each task.

There was a significant main effect of Tool (F(1, 15) = 90.28, $p < 0.001$). On average, participants spent 61.5 s longer on the task when they used DevTools, compared to when they used Poirot. There was also a significant main effect of Strategy (F(1, 15) = 26.29, $p < 0.001$). On average, participants spent 35.0 s longer on Custom tasks compared to F&R tasks. In addition, there was a significant Tool-by-Strategy interaction (F(1, 15) = 32.96, $p < 0.001$). The time saved by using Poirot (versus DevTools) was 78.4 s more on Custom tasks, compared to F&R tasks. Specifically, simple effect analysis showed that, on F&R tasks, participants spent 22.3 s longer when they used DevTools, compared to when they used Poirot (F(1, 15) = 6.35, $p < 0.05$). On Custom tasks, participants spent 100.7 s longer when they used DevTools, compared to when they used Poirot (F(1, 15) = 102.8, $p < 0.001$). In other words, while there was advantage in using Poirot overall, this was much larger on Custom tasks.

A second Tool by Strategy repeated measures ANOVA was conducted on the time participants spent on solving each task, however, tasks that participants failed to complete were excluded from the analysis. Overall, the results were consistent with the analysis that included all tasks, but the effects appeared to be slightly weaker (Fig. 7).

There was a significant main effect of Tool (F(1, 6.64) = 9.64, $p < 0.05$). On average, participants spent 12.1 s longer to find a solution to the task when they used DevTools, compared to when they used Poirot. There was also a marginal main effect of Strategy (F(1, 5.61) = 4.05, $p = 0.09$). In addition, there was a marginal Tool-by-Strategy interaction (F(1, 5.69) = 5.50, $p = 0.06$). The time saved by using Poirot (versus DevTools) was 35.0 s more on Custom tasks, compared to F&R tasks. Specifically, simple effect analysis showed that, on F&R tasks, there was no significant difference in time participants spent to find a solution (F(1, 14.66) = 2.54, $p = 0.13$). On Custom tasks, participants spent significantly more time when they used DevTools to find a solution, compared to when they used Poirot (F(1, 5.34) = 8.10, $p = 0.033$).

6.3 NASA TLX

Five of the six NASA TLX categories were used to determine a perceived cognitive load for each task: effort, frustration, mental demand, temporal demand, and performance. Each was rated on a 7-point Likert scale where 1 = Very Low and 7 = Very High. For the performance metric, the labels were adapted to 1 = Perfect and 7 = Failure. Perceived cognitive task load was calculated by averaging the individual scores from each category. Participants reported an average perceived cognitive load of 2.4 (sd = 0.8) when using Poirot compared to 4.6 (sd = 1.1) when using DevTools. A Wilcoxon-Pratt Signed Ranks test indicates that the difference in perceived cognitive load is significant (Z = 3.39, $p < 0.001$).

The following are all reported on a 5-point Likert scale. On a scale of “Not very often” to “Very often”, participants rated how often they used Chrome Developer Tools in their design work a 2.8 (sd = 1.2). On the same scale, they rated how often they would use Poirot in their design work a 3.9 (sd = 1.1). A Wilcoxon-Pratt Signed Ranks test indicates that the difference between the two ratings is significant (Z = −2.26, $p < 0.05$). On a scale from “Not very useful” to “Very useful”, participants rated the usefulness of the built in design system in Poirot a 4.5 (sd = 0.7). On the same scale, they rated the usefulness of the Poirot user interface a 4.1 (sd = 0.8). On a scale from “Not very intuitive” to “Very intuitive”, they rated the intuitiveness of the Poiret user interface a 4.1 (sd = 0.8). Finally, on a scale of “Not very representative” to “Very representative”, participants rated how representative the study tasks were of actual tasks they might attempt to do in Chrome Developer Tools a 3.9 (sd = 1.1).

In the post-survey many participants expressed interest in using Poirot for their work:

> Finally, a direct manipulation interface for designers to go in and tweak code with! - P1

> Overall, I think this tool is perfect. Just some small refines for the details, it would be very popular tool. - P13

Thirteen of the 16 users in their post survey left comments in their post-survey saying they found Poirot “simple”, “intuitive”, or “easy” to use:

> [I]t’s very easy to use, I can learn how to use this in one second. And this tool gives me a feeling that I’m using a design software like Sketch to modify the page in a familiar way. - P13

> Easy to use, similar to other design tools i am familiar with - P15

7 Discussion

The probability of whether a participant would successfully complete a design task in the user study was highly dependent on two factors: the tool being used and the strategy that could be employed to solve it. When a participant was using Poirot, there was a very high probability of successful completion of any task. In fact, participants using Poirot were successful in almost all tasks, only failing 4 of 144 tasks across all participants. In contrast, when using DevTools, participants collectively failed 53 tasks, and the probability of successfully completing a task was highly dependent on whether or not the task could be solved using a Find and Replace strategy.

As an example, task #2, update the background color for all gray800 buttons to blue500, was successfully completed by all but one of the participants. When participants selected one of the gray800 buttons, most quickly discovered a CSS rule with the selector “.primary-btn” and declaration “background-color: #333;”. To successfully solve this task, they changed #333 to #01a1dd, and all gray800 buttons updated to blue500.

In contrast, task #1, update the color of a single heading from gray800 to blue500, was only completed by five of the participants when using DevTools. When participants selected the heading, a CSS rule with a selector targeting just this element did not exist. Instead, there was simply a rule that targeted all headings throughout the page with the declaration “color: #333;”. Most participants changed this to #01a1dd, but after making the change, they discovered that all headings throughout the page were now blue500.

Poirot reduced the confusion in tasks such as these by following good design principles including direct manipulation (Shneiderman 1983), a clear conceptual model, and visibility (Norman 2013). Other confusion it eliminated entirely by hiding unneeded details.

For example, Poirot made it easy for users to select an element and clear which elements were going to be updated. In contrast, users struggled to select elements in

DevTools due to its indirect method for selecting elements. DevTools does support a direct manipulation method, but this is hard to find and was only used by one participant.

> [Poirot] highlighted on screen what I was looking at and made it a lot more intuitive on what I was editing and making changes to. I think it's hard with Chrome/Safari developer tools to really know what you're looking at with 100% accuracy in a short amount of time unless you're really an experienced coder. - P6

Following the design principle of consistency, Poirot's interface always showed a consistent UI, regardless of the underlying CSS declarations. In contrast, DevTools required designers to scroll through a list of CSS rules searching for the one that had the declaration property they were seeking.

Participants also struggled with knowing the exact syntax required and made syntax errors even when using correct declarations. These issues have been well documented (Park et al. 2013b, 2015), and our observations support their findings. Poirot hides this complexity of CSS syntax by removing it entirely.

In general, DevTools required four steps for completing a design task: selecting an element, finding where to modify the element, remembering CSS syntax, and evaluating the change to an element. Each of these steps presented complications to designers, either due to the difficulties of programming syntax or unfamiliar UI. Poirot assisted users through these complications by hiding complexity and the use of intuitive or familiar interfaces and interactions.

This paper adds understanding to the complicated roles of UI designers as web developers and the opportunity for improved communication with programmers (Putnam et al. 2016a). Poirot is one attempt to address pain points we discovered during our formative investigation, but there remains more opportunity here. It also sheds light on designers' challenges with current web inspector tools. Though we specifically studied designers, many insights should generalize to novice developers of varying backgrounds. It also presents several UI solutions including copying and pasting styles across websites, auditing a website's use of a design system, and visibility and control over multiple or single element selection.

8 Limitations and Future Work

In its current state, Poirot cannot persist changes across site code updates that change the underlying HTML structure. Strategies for identifying consistent elements across HTML structure changes could be employed to create a system more robust to changes (Chakrabarti et al. 2008; Cai et al. 2003). Poirot handled scores of websites we tested it on including Google, Amazon, and Wikipedia, but it wasn't able to prevent Javascript click events on sites such as Twitter or YouTube, precluding most items from being selected on those pages.

In the future, Poirot could be extended in numerous ways, such as design system component library support, automatically extracting design systems from existing websites, better element positioning support, animation and interaction support, as well as a list of minor improvements suggested by participants during the study. Because Poirot reminded participants of their traditional graphics tools, they expected it to support all the features they are used to such as undo/redo, drag to reposition panels, and shift to multiselect.

Regardless of any limitations, the concepts upon which Poirot is built were shown to be helpful to designers and provide valuable insights into future tools. Designers were faster and more successful in making changes when using Poirot, and they perceived Poirot to be easier to use than DevTools. By creating tools that leverage the expertise of individuals rather than force them to adapt to programmers' tools, the diversity of individuals making valuable contributions to websites can increase. The same principles that helped designers improve the aesthetic quality of a website could be applied to making a website more accessible or inclusive. The benefit of these edits would be enormous if they could be crowdsourced and shared with others.

9 Conclusion

In the modern ecosystem, designers must often rely heavily on developers to visually update websites. To understand these pain points, we conducted interviews with professional designers and developers, surveyed professional UI/UX designers, and collected and analyzed professional design documents. This investigation provided insights that informed the design and implementation of Poirot, a web inspection tool for designers. We evaluated Poirot with professional UI/UX designers in a within-subjects lab study, and discovered that compared to Chrome DevTools, participants completed more tasks with Poirot and were enthusiastic about the prospect of such a tool to assist in their design work. We concluded with a discussion of the difficulties we observed designers experience when working with a popular web inspection tool. We hope to see these findings inform the creation of better web tools for designers and improve the communication between designers and developers.

References

Alimadadi, S., Sequeira, S., Mesbah, A., & Pattabiraman, K. (2014). Understanding JavaScript event-based interactions. In *Proceedings of the 36th International Conference on Software Engineering* (*ICSE 2014*) (pp. 367–377). ACM, New York, NY, USA. https://doi.org/10.1145/2568225.2568268.

Amiri, F. (2011). Programming as design: The role of programming in interactive media curriculum in art and design. *International Journal of Art & Design Education*, *30*(2), 200–210.

Blackwell, A. F. (2002). First steps in programming: A rationale for attention investment models. In *Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments*. IEEE, 2.

Bolin, M., Webber, M., Rha, P., Wilson, T., & Miller, R. C. (2005). Automation and customization of rendered web pages. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology* (*UIST '05*) (pp. 163–172). ACM, New York, NY, USA. https://doi.org/10.1145/1095034.1095062.

Burg, B., Ko, A. J., & Ernst, M. D. (2015a). Explaining visual changes in web interfaces. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (pp. 259–268). ACM.

Burg, B., Ko, A. J., & Ernst, M. D. (2015b). Explaining visual changes in web interfaces. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (*UIST '15*) (pp. 259–268). ACM, New York, NY, USA. https://doi.org/10.1145/2807442.2807473.

Cai, D., Yu, S., Wen, J.-R., & Ma, W.-Y. (2003). Vips: a vision-based page segmentation algorithm.

Chakrabarti, D., Kumar, R., & Punera, K. (2008). A graph-theoretic approach to webpage segmentation. In *Proceedings of the 17th international conference on World Wide Web* (pp. 377–386). ACM.

Chang, K. S.-P. & Myers, B. A. (2012a). WebCrystal: Understanding and reusing examples in web authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3205–3214). ACM.

Chang, K. S.-P. & Myers, B. A. (2012b). WebCrystal: Understanding and Reusing Examples in Web Authoring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '12*) (pp. 3205–3214). ACM, New York, NY, USA. https://doi.org/10.1145/2207676.2208740.

Chilana, P. K., Alcock, C., Dembla, S., Ho, A., Hurst, A., Armstrong, B., & Guo, P. J. (2015). Perceptions of non-CS majors in intro programming: The rise of the conversational programmer. In *2015 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 251–259). IEEE.

Dorn, B., & Guzdial, M. (2006). Graphic designers who program as informal computer science learners. In *Proceedings of the second international workshop on Computing education research* (pp. 127–134). ACM.

Dorn, B., & Guzdial, M. (2010a). Discovering computing: perspectives of web designers. In *Proceedings of the Sixth international workshop on Computing education research* (pp. 23–30). ACM.

Dorn, B., & Guzdial, M. (2010b). Learning on the job: characterizing the programming knowledge and learning strategies of web designers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 703–712). ACM.

Fitzgerald, M. J. et al. (2008). *CopyStyler: Web design by example*. Ph.D. Dissertation. Massachusetts Institute of Technology.

Frost, B. (2016). Atomic design.

Fujima, J., Lunzer, A., Hornbæk, K., & Tanaka, Y. (2004). Clip, connect, clone: Combining application elements to build custom interfaces for information access. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (*UIST '04*) (pp. 175–184). ACM, New York, NY, USA. https://doi.org/10.1145/1029632.1029664.

Goodman, E., Stolterman, E., & Wakkary, R. (2011). Understanding interaction design practices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '11*) (pp. 1061–1070). ACM, New York, NY, USA. https://doi.org/10.1145/1978942.1979100.

Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communication ACM*, *28*(3), 300–311. https://doi.org/10.1145/3166.3170.

Head, A., & Hearst, M. A. (n. d.). Tutorons: Generating context-relevant, on-demand explanations and demonstrations of online code. ([n. d.]).

Herring, S. R., Chang, C.-C., Krantzler, J., & Bailey, B. P. (2009). Getting inspired!: Understanding how and why examples are used in creative design practice. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 87–96). ACM.

Hibschman, J., & Zhang, H. (2015). Unravel: Rapid web application reverse engineering via interaction recording, source tracing, and library detection. In *Proceedings of the 28th Annual ACM*

Symposium on User Interface Software & Technology (*UIST '15*) (pp. 270–279). ACM, New York, NY, USA. https://doi.org/10.1145/2807442.2807468.

Hibschman, J., & Zhang, H. (2016a). Telescope: Fine-tuned discovery of interactive web UI feature implementation. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (pp. 233–245). ACM.

Hibschman, J., & Zhang, H. (2016b). Telescope: Fine-tuned discovery of interactive web UI feature implementation. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (*UIST '16*) (pp. 233–245). ACM, New York, NY, USA. https://doi.org/10.1145/2984511.2984570.

Hong, J., & Wong, J. (2006). Marmite: End-user programming for the web (CHI '06) (pp. 1541–1546). *ACM*.

Ko, A. J., Myers, B. A., & Aung, H. H. (2004). Six learning barriers in end-user programming systems. In *2004 IEEE Symposium on Visual Languages-Human Centric Computing* (pp. 199–206). IEEE.

Lieber, T., Brandt, J. R., & Miller, R. C. (2014). Addressing misconceptions about code with always-on programming visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '14*) (pp. 2481–2490). ACM, New York, NY, USA. https://doi.org/10.1145/2556288.2557409.

Lim, S. (2017). Ply: Visual regression pruning for web design source inspection. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (*CHI EA '17*) (pp. 130–135). ACM, New York, NY, USA. https://doi.org/10.1145/3027063.3048427.

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Constellation.

Oney, S., & Myers, B. (2009). FireCrystal: Understanding interactive behaviors in dynamic web pages. In *2009 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 105–108). https://doi.org/10.1109/VLHCC.2009.5295287.

Park, T. H., Dorn, B., & Forte, A. (2015). An analysis of HTML and CSS syntax errors in a web development course. *ACM Transactions on Computing Education (TOCE)*, *15*(1), 4.

Park, T. H., Saxena, A., Jagannath, S., Wiedenbeck, S., & Forte, A. (2013a). OpenHTML: Designing a transitional web editor for novices. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (*CHI EA '13*) (pp. 1863–1868). ACM, New York, NY, USA. https://doi.org/10.1145/2468356.2468690.

Park, T. H., Saxena, A., Jagannath, S., Wiedenbeck, S., & Forte, A. (2013b). Towards a taxonomy of errors in HTML and CSS. In *Proceedings of the Ninth Annual International ACM Conference on International Computing Education Research* (*ICER '13*) (pp. 75–82). ACM, New York, NY, USA. https://doi.org/10.1145/2493394.2493405.

Putnam, C., Reiner, A., Ryou, E., Caputo, M., Cheng, J., Allen, M., et al. (2016a). Human-centered design in practice: Roles, definitions, and communication. *Journal of Technical Writing and Communication*, *46*(4), 446–470.

Putnam, C., Reiner, A., Ryou, E., Caputo, M., Cheng, J., Allen, M., et al. (2016b). Human-centered design in practice: Roles, definitions, and communication. *Journal of Technical Writing and Communication*, *46*. https://doi.org/10.1177/0047281616653491.

Rode, J., Howarth, J., Pérez-Quiñones, M. A., & Rosson, M. B. (2005). An end-user development perspective on state-of-the-art web development tools.

Rode, J., & Rosson, M. B. (2003). Programming at runtime: Requirements amp; paradigms for nonprogrammer web application development. In *Proceedings of IEEE Symposium on Human Centric Computing Languages and Environments, 2003* (pp. 23–30). https://doi.org/10.1109/HCC.2003.1260198.

Rosson, M. B., Ballin, J., & Rode, J. (2005). Who, what, and how: A survey of informal and professional web developers. In *2005 IEEE Symposium on Visual Languages and Human-Centric Computing* (pp. 199–206). IEEE.

Shneiderman, B. (1983). Direct manipulation: A step beyond programming languages. *Computer*, *8*, 57–69.

Wang, Y. D., & Zahadat, N. (2009). Teaching web development in the web 2.0 era. In *Proceedings of the 10th ACM conference on SIG-information technology education* (pp. 80–86). ACM.
Wolf, T. V., Rode, J. A., Sussman, J., & Kellogg, W. A. (2006). Dispelling "Design" as the black art of CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (*CHI '06*) (pp. 521–530). ACM, New York, NY, USA. https://doi.org/10.1145/1124772.1124853.

Applying Design Thinking Practices

Getting Hands-on with Tele-Board MED: Experiencing Computer-Supported Teamwork in Therapist-Patient Sessions

Anja Perlich, Miriam Steckl, Julia von Thienen, Matthias Wenzel and Christoph Meinel

Abstract Doctor-patient sessions require more than the mere application of the doctor's medical knowledge with respect to the patient. Correct diagnoses and effective treatments depend strongly on a functioning interaction between doctor and patient, as well as on high-quality case documentation. We develop the software system Tele-Board MED (TBM), which offers unique support for medical consultations by allowing doctors and patients to take digital notes jointly. This chapter describes the hands-on experience psychotherapists make when using TBM for the first time in consultation sessions with patients. We look at three interlinked aspects of computer-supported therapist-patient teamwork: (i) the therapists' user experience regarding TBM, (ii) the interaction between therapist, patient and the TBM system and (iii) the effectiveness of a TBM feature to generate official clinical documents automatically. The study shows that even in the very first treatment session with TBM, therapists come to feel comfortable taking open, digital notes. TBM is used by therapists and patients not only for documentary purposes, but also as a tool to facilitate the therapeutic conversation. Regarding the therapist's administrative task of writing official clinical case reports, the study shows that even therapists who use TBM for the first time save 60% of their regular working hours when compiling official clinical documents after treatment sessions.

A. Perlich (✉) · M. Steckl · J. von Thienen · M. Wenzel · C. Meinel
Hasso Plattner Institute for Digital Engineering, Prof.-Dr.-Helmert Straße 2-3, 14482 Potsdam, Germany
e-mail: anja.perlich@hpi.de

M. Steckl
e-mail: miriam.steckl@hpi-academy.de

J. von Thienen
e-mail: julia.vonthienen@hpi.de

M. Wenzel
e-mail: matthias.wenzel@hpi.de

C. Meinel
e-mail: meinel@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_15

1 Introduction

When doctors treat patients, they need to handle several activities as once, which go beyond the mere application of medical knowledge. Being an agreeable and trustworthy conversation partner is an important part of the care doctors give to their patients. Furthermore, healthcare providers have to fulfill the professional duties of maintaining records of their patient cases. By law, doctors need to document their treatment promptly and comprehensively in a patient record (Bundesgesetz 2013). The legal duty of documentation also includes providing patients with an electronic copy of the treatment notes when requested. For doctors it can be demanding to be both a trustworthy conversation partner and a diligent document-keeper. This challenge is especially pronounced in the domain of behavior psychotherapy, where patients come with mental health problems, such as phobias, personality disorders and depression. In behavior psychotherapy, the relationship with the patient—that is greatly determined by the quality of consultation conversation—is a major predictive factor for treatment success (Grawe 2005). At the same time, it is important for therapists to take notes and create memory aids. In earlier studies, we learned that behavior psychotherapists take handwritten notes during or after the patient sessions to recall the session conversation (Perlich and Meinel 2015). Furthermore, note-taking in psychotherapy supports administrative processes, e.g. the application to health insurance companies for treatment funding. Writing official clinical documents, such as the case report for the health insurance is a very time-consuming and undesirable activity because therapists need to transfer their handwritten notes into a digital format (von Thienen et al. 2015).

We are rethinking note-taking in mental healthcare with the aim of smoothly integrating the processes of patient treatment and case documentation. Our goal is to transform documentation from a necessity, carried out by the care provider, into a beneficial activity that involves, engages and empowers the patient. We develop the Tele-Board MED (TBM) software system, which offers unique support for patient-doctor interactions by allowing doctors and patients to take digital notes jointly (Fig. 1). Furthermore, TBM offers a report generation feature. This allows the creation of official clinical documents in an automated way.

With the usage of Tele-Board MED in medical consultations, we follow three goals:

(1) Supporting faster, high-quality documentation,
(2) Supporting patient-doctor teamwork,
(3) Providing a positive learning experience for users.

In earlier studies, we learned that TBM has great potential for making documentation more efficient (von Thienen et al. 2015) and that it supports the doctor-patient relationship and the empowerment of patients (Perlich and Meinel 2016). However, we also learned that the therapist's perception of the effort involved in learning TBM is higher than their perception of the benefits (Perlich et al. 2018).

This chapter describes the hands-on experience psychotherapists make when using TBM for the very first time in patient interviews. We conducted a user experi-

Fig. 1 Tele-Board MED user scenario: During a treatment session, patient and therapist take digital notes jointly. The interactive whiteboard screen shows a documentation panel with sticky notes, which contain key phrases captured during the conversation. The therapist uses a keyboard as one possible means to enter information. At the end of the session, the patient can receive a printout of the session notes to take home

ence study with behavior psychotherapists focusing on three interlinked aspects of computer-supported therapist-patient teamwork: (i) the therapists' user experience regarding TBM, (ii) the interaction between therapist, patient and the TBM system, and (iii) the effectiveness of a TBM feature to generate official clinical documents automatically. Part 2 of this chapter outlines the Tele-Board MED system and its usage for digital note-taking in doctor-patient sessions and for the automated creation of medical reports. Part 3 provides a detailed review of a user experience study. This includes a description of study goals, the study design, spanning the preparatory sessions with participants, the execution of therapist-patient interviews supported by TBM, the therapists' evaluation of TBM usage as well as video-based analyses of therapist-patient interactions. The chapter will close with a conclusion in part 4.

2 The Tele-Board MED System

Tele-Board MED (TBM) is a tool that allows doctors and patients to create electronic patient files jointly during treatment sessions. Figure 1 shows a treatment session scenario supported with TBM. Doctor and patient can freely edit and fill TBM's whiteboard interface with sticky notes, scribbles, and pictures. TMB builds on the Tele-Board system, which was designed for supporting creative teamwork over distances (Gumienny et al. 2011). Next to the whiteboard interface for collaborative documentation, TBM offers a medical report generation feature to create clinical documents out of the digital notes taken during the session. Thus, doctors need to capture information only once and can make multiple use of it.

The potential of computer support for psychological wellbeing has been recognized already more than a decade ago (Coyle et al. 2007). There is a remarkable amount of work on the computerization of therapy, i.e. technology-supported alternatives to consultations in a doctor's office. The available offers range from anonymous online counselling to guided self-help applications, for specific mental health problems, all the way to relaxation and meditation apps. However, the number of tools designed as an integrable extension to face-to-face treatment sessions is rather sparse (Knowles et al. 2014). One example of this so-called computer-mediated therapy is a computer game designed to be played by patients together with their therapists (Coyle et al. 2007). To our knowledge, there is no digital tool to facilitate the sharing of mental health record notes with patients. However, the topic of patient access to electronic health records in mental healthcare is actively discussed. Kahn et al. (2014) are advocates of this approach and argue that showing patients their mental health records can lead them to more active interest in their health and to increased engagement in their treatment process. With TBM, collaborative note-taking in doctor-patient sessions is made possible. Moreover, it supports the reutilization of digital notes for the automatic generation of medical reports.

2.1 Digital Note-Taking with Tele-Board MED

Since TBM is a web-based system it can be used on various devices—from stationary hardware, such as a digital whiteboard or desktop computer, to mobile devices, such as laptop, tablet computer or smartphone. On a big touch screen, doctor and patient can work on the whiteboard panel in a collaborative manner (cf. Fig. 1). They can document on a blank panel or work with templates for specific purposes, topics and exercises. TBM provides multiple ways of data entry through typing, handwriting and speech recognition (Wenzel et al. 2019). In the user experience study presented in this chapter, we provided a wireless keyboard and a tablet computer in addition to the digital whiteboard (cf. Fig. 2). Digital notes can be captured by typing on a keyboard, which also contains a trackpad for moving the mouse pointer. As an alternative to keyboard typing, TBM provides a sticky pad app. Just like writing on a paper sticky note pad, the app allows writing notes with a digital pen on a mobile device. The handwriting recognition feature transforms scribbles into computer text. A press of a button sends the sticky note to the whiteboard interface.

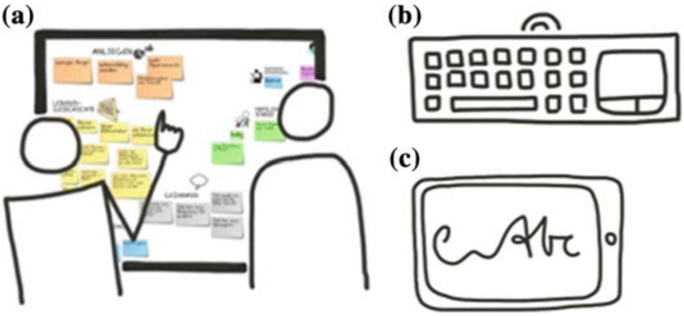


Fig. 2 Possible input devices in a Tele-Board MED user scenario: **a** digital whiteboard with **b** a wireless keyboard with trackpad and **c** tablet computer running the TBM sticky pad app with handwriting and speech recognition

2.2 Creating Medical Reports with Tele-Board MED

Medical reports are a broadly used means of medical documentation and communication between healthcare stakeholders. They provide well-organized summaries of case details including symptoms and diagnoses, treatment plans and outcomes. For care providers, they serve as a central tool to exchange case information and to request or justify treatment remuneration. The content and wording of medical reports is highly important. Every single mistake can have serious consequences. When care providers base their treatment decisions on improper case information they may schedule faulty interventions. Furthermore, patients can sue their care providers for an inappropriate diagnosis or treatment.

A central document in the psychotherapy domain is the case report to health insurance companies, which is necessary to acquire funding for the patient treatment. Currently, the creation of case reports is very inefficient, because therapists document treatment sessions by hand first and retype the information later to create digital reports (von Thienen et al. 2015). The written case reports are condensed summaries of the information addressed in the therapy sessions and contain a case description, a problem analysis and a treatment plan. The German National Association of Statutory Health Insurance Physicians provides guidelines on case report creation (Bundesvereinigung 2017). Thus, there is potential to support the report creation with automation by reusing digital session notes. Within TBM, we implemented the medical report generation feature, which allows sorting the whiteboard panel content into the corresponding sections of a text file automatically (cf. Fig. 3).

We prepared a collection of whiteboard templates that reflect prominent topics in early therapeutic patient interviews and, at the same time, cover all relevant information for the case report. There are seven whiteboard panels for the following topics: concerns and symptoms, patient history, behavior analysis, therapy plan, psychological finding, somatic finding and diagnosis. The panels contain headlines, visuals and prepared sticky notes to be filled in. Thus, the panels serves as an aid to the conversation, offering the function of an interview guideline and a framework for

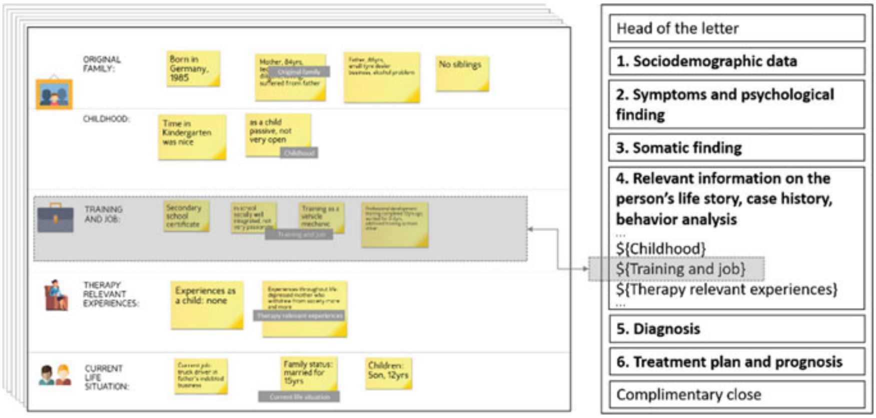


Fig. 3 Concept of the report generation feature. Based on their spatial arrangement, the whiteboard panel content (left) is sorted automatically into the sections of a text file (basic structure of a report template on the right). The two grey rectangles and the connecting arrow indicate that the text on sticky notes positioned in the central portion of the whiteboard panel will appear in the section “Training and job” of the automatically created case report

capturing session notes. On top of this whiteboard layer is a second layer dedicated to report creation, linking certain panel areas to the sections of the report. Figure 3 shows a whiteboard panel with information about the patient's life history (left) and a schematic diagram of the case report template (right). We created a case report template in .docx format with sections based on the guidelines for case report creation. This text document contains a header, six section headings and a footer with complementary close. Below the section headings the document contains placeholders characterized by a dollar sign and curly brackets (e.g. ‘${Training and job}’). When creating the case report, the placeholders are replaced by the sticky note texts. For example, the placeholder ‘${Training and job}’ will be replaced with the text describing the patient's life history regarding school, training and professional development (cf. Fig. 2). Equipped with these augmented whiteboard panels, creating sticky notes on the fly during the treatment session already prepares the report creation in the background. After the patient interview, the therapist just needs to click a button and the case report is created automatically.

In case medical reports were needed that followed a different scheme (e.g., medical discharge letters in a hospital with specific topic requests or medical reports in a domain other than psychotherapy), TBM templates can easily be adjusted. The automatic report creation feature is highly flexible in terms of content.

3 User Experience Study with Therapists

In order to find out whether the TBM system can keep its promises on collaborative note-taking and automatic medical report creation, we tested the system with behavior psychotherapists in a hands-on user experience study (Perlich and Meinel 2018). Psychotherapy is a very sensitive domain. Therapists commonly seek to avoid all degrees of uncertainty in the treatment process, which novel tools could create from a therapist's point of view. Thus, we designed a study, which involves simulated instead of real patients. The core of the study is a 50-min anamnesis session dialogue between a behavior psychotherapist and a patient actor, who presents a particular mental health problem. They used TBM for collaborative note-taking and the session was video-recorded. After the anamnesis interview, we evaluated the therapists' user experience and had the TBM report generation feature tested. Following up on the user experience session, we analyzed the recorded video material of the anamnesis sessions (cf. Fig. 4).

3.1 Study Participants and Setup

We conducted the study in Germany with four behavior psychotherapists who are either practicing or in training to become approved therapists. Table 1 shows an overview of the participants including sociodemographic data, therapy experience and report writing habits. Furthermore, we included four volunteers who acted as patients. In order to acquire volunteers for acting out the patient role we asked colleagues from the Hasso-Plattner-Institute in Potsdam, Germany. They were asked to memorize the description of a realistic clinical case, including biographical data and a certain mental health problem including symptoms and experienced unpleasant situations. Two female volunteers acted out the role of a 32-year old woman suffering from a social phobia and two male volunteers played a 27-year old man suffering from an obsessive-compulsive disorder. Both patient case descriptions stem from Reinecker's educational book on clinical psychology (1999). To keep the anamnesis interview as authentic as possible the therapists were unfamiliar with the patient actors and the case they would present. The therapists were prepared in an individual, moderated, 2-h introductory session where they tried out the TBM whiteboard inter-

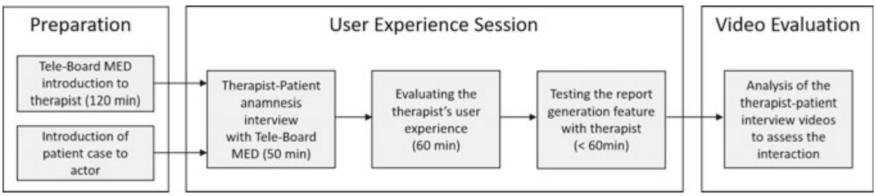


Fig. 4 The course of the Tele-Board MED user experience study

Table 1 Characteristics of the behavior psychotherapists who participated in the study (N = 4)

Information about therapists				
#	1	2	3	4
Gender	Male	Female	Female	Female
Age	26	32	27	27
Therapy experience	Therapist in training year 2	Practicing therapist over 2 years	Therapist in training year 1	Therapist in training year 2
Number of written case reports	≥20	≥20	6–10	11–20
Average time per case report (min)	270	150	240	270
Documentation with TBM: preferred note-taking device	Keyboard	Keyboard	Keyboard	Tablet with handwriting recognition
Documentation with TBM: preferred navigation device	Keyboard-integrated trackpad	Keyboard-integrated trackpad	Whiteboard touch interface	Whiteboard touch interface
Report creation with TBM: time needed for revision (min)	40	30	42	33
Report creation with TBM: estimated further time needed (min)	60	30	50	90

face, the prepared documentation templates (e.g., patient's life history, cf. Fig. 3, left side), and the report generation feature. They adapted the documentation templates based on their personal preferences to be well prepared for the anamnesis interview. Furthermore, they were invited to try out the TBM sticky pad app and its handwriting recognition feature.

After the recruitment and preparatory session with study participants, we scheduled a date for the live treatment session, in which the therapist and simulated patient conducted an anamnesis interview with TBM. The session took place in a meeting room, which was refurnished to resemble a therapy room and equipped with TBM devices (Fig. 5). Therapist and patient were sitting facing each other and the digital whiteboard. On a small side table, the wireless keyboard and the tablet computer

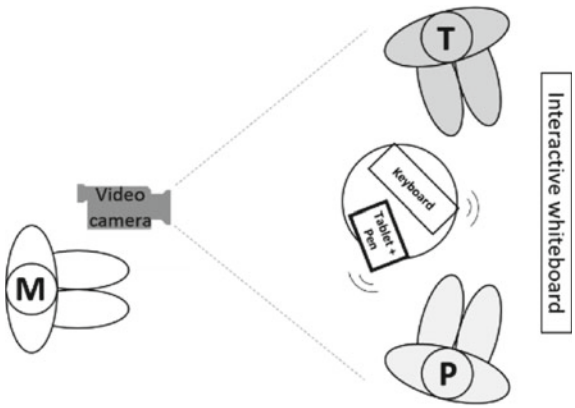


Fig. 5 Spatial setup for the therapist-patient interview in the user experience study. Therapist (T) and patient (P) sit in front of the interactive whiteboard. On a side table, tablet computer and wireless keyboard are available. The moderator (M) captures the conversation with a video camera

with digital pen were available as input devices. Additionally, a digital video camera was mounted on a tripod to record the therapist-patient session.

3.2 Goals and Hypotheses

In the user experience study, we test the following hypotheses related to the three goals of Tele-Board MED (TBM).

(1) **Supporting faster, high-quality documentation**. TBM shall support the in-session note-taking and administrative documentation tasks. With TBM, we aim to accelerate the process of creating official clinical documents by automatically reusing digital session notes. We are testing the following hypothesis based on the evaluation of the TBM report generation feature:

    H1 Creating case reports with TBM takes less time than therapists usually spend on writing case reports to the health insurance.

(2) **Supporting patient-doctor teamwork**. The benefits of TBM shall go beyond digital documentation. TBM shall facilitate a lively exchange in patient-doctor conversations. We are testing the following hypothesis based on the video analysis:

    H2 Therapists and patients do not only use TBM for the entry of information, but also as a tool to facilitate the therapeutic conversation—as reflected by pointing gestures to the digital whiteboard that are statistically unrelated to note-taking activities.

(3) **Providing a positive learning experience for users**. TBM shall provide a positive experience for both the patient and therapist user. In this study, we focus on the therapists as the primary hands-on user, because they are a step ahead of the patients regarding the familiarity with the system. We are going to

test the following hypotheses based on the video analysis and the therapist user experience evaluation:

H3.1 Within one session, the therapists attain a feeling of satisfaction in the use of TBM.
H3.2 Within one session, there is a marked trend towards fluent documentation with TBM—as reflected by a decrease of data input problems.

3.3 Evaluating the Report Generation Feature

We hypothesized that creating case reports with TBM takes less time than therapists traditionally spend on writing case reports (hypothesis H1). In order to assess the TBM report generation feature we asked the therapists to create a case report based on the patient session. It took them two clicks and a few seconds waiting time until TBM generated the initial version of the case report. We asked them to revise it as if they wanted to submit it to the health insurance company. While the notes on the whiteboard are in the form of key points in colloquial language, the case reports are written in a professional jargon and partly in subjective tense. Thus, therapists revised the generated text document by formulating sentences, paraphrasing notes and changing the inflection of words. When the therapists' perception differed from the patient-reported information, they supplemented the information. At times, text was also removed or added when it was not covered in the session notes. Some therapists highlighted text passages, which still needed additional information. All four therapists created an intermediate draft and considered collecting additional information in a follow-up session. It is a common procedure to schedule up to five trial sessions after the initial anamnesis interview until therapists send out the report. We recorded the time they needed for the revision of the generated report. Furthermore, we asked them via a questionnaire to estimate the additional time they would need to finish the report and how much time they usually spend on writing a case report.

Results: Table 1 shows information about the therapists and the report creation variables. For example, therapist 2, who usually needs an average of 150 min to write a case report, needed 30 min to revise the automatically generated report. She expected an additional 30 min to finalize the report with the information collected in follow-up sessions. On the average, it took the therapists 36 min to revise the report generated by TBM. After the revision, they expected it to take an average of 56 min to complete the report. With their current approaches to case report creation the therapists need an average of 3.9 h (233 min) per case report. This means that the report creation time is reduced by 60%—even in scenarios where therapists use TBM for the very first time.

3.4 Tele-Board MED User Experience Evaluation

Directly after the anamnesis interview, we sent off the patient and conducted a user experience evaluation with the therapist. They were asked to express their feeling over the 50-min period in the form of a hand-drawn and annotated user experience curve (Kujala et al. 2011). We hypothesized that within one session, the therapists attain a feeling of satisfaction in the use of TBM (hypothesis H3.1).

Results: When we look at the user experience (UX) curves of the four therapists, we see a prominent improvement in the level of satisfaction. Figure 6 shows a combined diagram of the UX curves of all four therapists. The comparison of curves suggests a pattern: during the first third, the curves of all four therapists start in the area of negative or neutral feelings. During the second third, the curves of therapists T1, T2 and T4 oscillate around the zero line back and forth, indicating shifts between slight satisfaction and slight dissatisfaction. In the final third of the treatment session, the curves of all therapists rise in the positive emotion range to a medium or high level.

Figure 6 also shows the annotations of therapist number 2, which relate to the highlighted curve. In the beginning, the therapist felt insecure, and she looked at the whiteboard often. Her feeling started to get better and better. Then some confusion came up when she wanted to add a line break on a sticky note. She pressed the Enter key, which leads to closing the sticky note editing mode. In the middle of the conversation, the therapist and patient jumped a lot between topics, accompanied by some switches between whiteboard panels. When working on one panel at a time (e.g., the patient's life history), the therapist achieved a secure feeling. Towards the

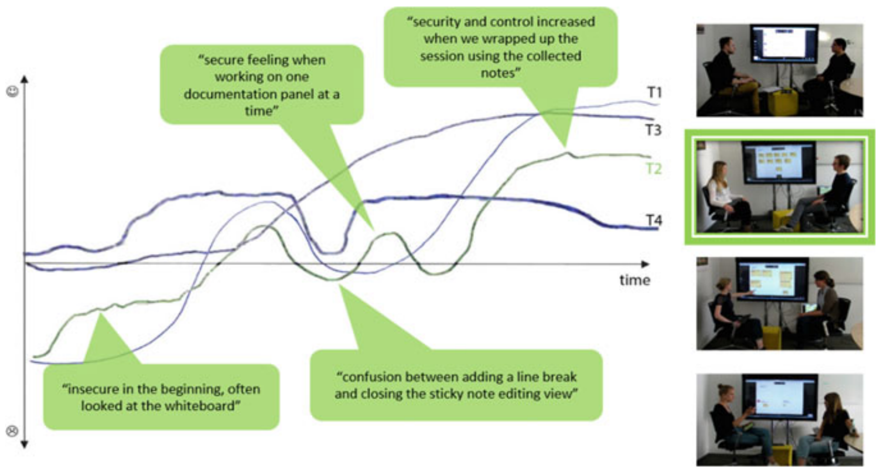


Fig. 6 Comparison of user experience curves drawn by the four therapists, which represent their satisfaction over time. The curve of therapist number 2 is highlighted and contains the therapist's written annotations. The thumbnail pictures on the right are screenshots of the four session recordings

end of the session, she felt increasingly secure and in control, also because she could use the collected notes to wrap up the session with the patient together.

The choice of input and navigation devices differed from therapist to therapist. Therapists 1, 2 and 3 chose the keyboard as the preferred input device, and therapist 4, by contrast, favored the tablet with digital pen and handwriting recognition. For navigating the screens or moving around sticky notes, therapists 1 and 2 used the trackpad mouse integrated into the keyboard. Therapists 3 and 4, on the other hand, decided to use the whiteboard touch feature for navigation (cf. Table 1).

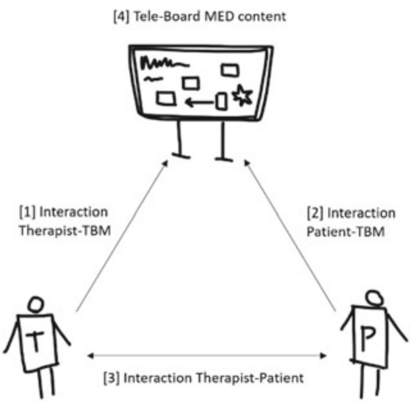
3.5 Assessing Interactions Through Video Analysis

We investigated the use of the TBM system as well as the behavior and interactions of the therapist and patient by conducting a video analysis.

In terms of related works, it can be noted that psychotherapy sessions are often video-recorded, because this is a useful means of assessing and improving the skills and abilities of therapists in training. By means of session video recordings, supervising therapists can support learners with feedback and advice (Topor 2017). In order to analyze qualitative data such as videos or audio recordings in a mathematical way, coding schemes are important tools. We looked at previous work on video content analysis and coding schemes for teamwork scenarios, especially in dyads. Bakeman and Adamson (1984) measured the interaction of infants with mothers and peers based on taped video cassettes. They developed a coding scheme that captured the different states of engagement of the infants, such as watching or interacting with their mothers or with objects. In the domain of couple therapy, Gottman and Levenson (2000) propose a coding scheme for video recordings to assess marital conflict discussions. Their Rapid Couples Interaction Scoring System (RCISS) contains categories divided into behaviors of the speaker and behaviors of the listener, which are coded on each speech turn taking. In the same study, another coding scheme for emotions was used in sequences where conflicts arose (Specific Affect Coding System, SPAFF). Peluso et al. (2018) applied the SPAFF to determine the quality of therapeutic relationships in counselling sessions. Jung (2011) modified the RCISS to apply it to a scenario of pair programming (i.e. two people working at one computer to produce software code in a collaborative manner) consisting of codes for positive and negative aspects in speaker and listener. The scenario of two people interacting with one computer is comparable to the TBM user scenario. However, Jung's coding scheme does not reflect the interaction with a computer system. Ramseyer and Tschacher (2011) performed an analysis of videos of psychotherapeutic consultation sessions with a special focus on the coordination of patient's and therapist's movement (so called synchrony) as one aspect of therapeutic alliance. They developed an automated objective video analysis algorithm (Motion Energy Analysis; MEA) in order to quantify nonverbal behavior in dyads.

The review of related work shows no coding scheme for analyzing the interaction of (therapist-patient) dyads with joint computer usage. Due to the lack of an existing theoretical model, we developed a coding scheme for human-human-computer interaction. We took an inductive approach and derived the coding scheme from the video material itself by investigating gestures, conversation and system operating modes. We developed a coding scheme with four main categories: (1) interaction therapist-TBM,, (2) interaction patient-TBM, (3) interaction therapist-patient, and (4) TBM content (cf. Fig. 7). The TBM content category encompasses codes for whiteboard template panels shown on the screen, which represent the conversation topics. The coding scheme with the four main categories and subordinate codes was refined in iterative cycles when watching the video. In the initial version, the subordinate codes for the TBM interaction therapist and patient were similar. However, the patients did not enter information through whiteboard, keyboard or tablet themselves and therefore the codes were not applied. The revision of the coding scheme also included the addition of three codes on data input problems when writing digital sticky notes. The table in Fig. 7 shows the final coding scheme with 27 codes in total. The codes cover all the actions, which took place during the four therapy sessions. Finally, all video sequences were allocated with one specific code for all of the four main categories. Thus, the subordinate codes of each of the four main categories are mutually exclusive and collectively exhaustive.

After setting up the coding scheme, the video material was coded with the software tool MAXQDA 2018 (Fig. 8). Altogether, 200 min of audiovisual material were analyzed (stemming from 50-min treatment sessions of four therapists) and a data set with 4309 codings was obtained. One main coder was in charge of coding the four



[1] Interaction Therapist-TBM
No interaction with TBM (full attention to patient)
Attention to whiteboard – eyes on board
Attention to whiteboard – finger pointing towards board
Attention to whiteboard – finger touch input
Keyboard input – typing text
Keyboard input – using trackpad
Tablet input without sending sticky note
Tablet input with sending sticky note (handwriting)
Tablet input with sending sticky note (computer text)
Input problem – change sticky note position or size on board
Input problem – add line break on sticky note
Input problem – start sticky note editing mode
[2] Interaction Patient-TBM
No interaction with TBM (full attention to therapist)
Attention to whiteboard – eyes on board
Attention to board – finger pointing towards board
[3] Interaction Therapist-Patient
Therapist speaking
Patient speaking
Silence
Silence because of interaction with TBM
[4] Tele-Board MED content
Concerns and symptoms
Diagnosis
Psychological finding
Somatic finding
Therapy plan
Behavior analysis
Patient history
Other

Fig. 7 Coding scheme for the qualitative video content analysis for assessing the interaction in psychotherapy sessions with Tele-Board MED (TBM)

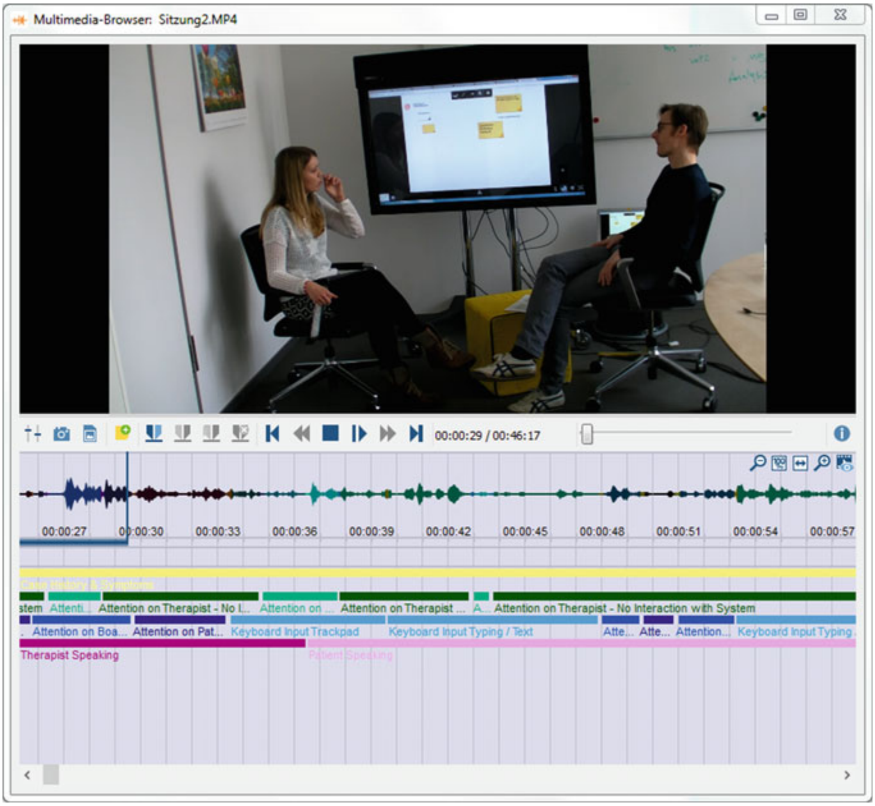


Fig. 8 Video coding with MAXQDA: the bars in the bottom represent the codings of the video sequences. There are four bars for the four main code categories. Each color represents a code

videos. A second coder was involved to double-check samples of video section codings. Capturing the beginning and ending of coded video sequences (e.g., "therapist enters text via the keyboard" minute 3:11 to 3:14) rendered the data appropriate for quantitative analyses. Based on the emotion curves drawn by the therapists, which had indicated different dynamics in the first, second and third part of the session (cf. Sect. 3.4), we also split the coding data set in three parts (each session part being 50/3 = 16.6 min long). Thus, coding frequencies could be compared over time. Mathematical analyses of coding frequencies were conducted with SPSS 25.

We hypothesized that therapists and patients do not only use TBM for the entry of information, but also as a tool to facilitate the therapeutic conversation—as reflected by pointing gestures to the screen that are statistically unrelated to note-taking activities (hypothesis H2). Pointing gestures seem a suitable metric to assess how therapist and patient make active use of the digital documentation content by referencing to the whiteboard screen apart from mere note-taking obligations. Generally, there is no significant correlation of pointing gestures and text input in any of the session parts, indicating that therapists and patients indeed reference the screen content in situations that are unrelated to note-taking. Thus, pointing gestures appear to be part of therapeutic conversations rather than being elicited by documentation obligations.

Results: Both therapists and patients spontaneously gesture towards the board during their discussions. Therapists use pointing gestures a little more often than patients, averaging 4.5 gestures per session part. During the middle session part, the number of pointing gestures is slightly reduced, averaging 3 gestures per therapist, whereas higher numbers of pointing gestures are observable amongst therapists in the first session part (5.75 on the average) and the last session part (4.75 on the average). Thus, the active use of the TBM templates and documented content can be observed, indicating that the system is not only used for the entry of information, but also as a tool to facilitate the conversation. Three out of four patients also use pointing gestures in the course of the conversation. Over time, there is a slight increase in pointing gestures among patients, which might reflect a process of patient empowerment in terms of an increasing active engagement with the record content. Only one patient gestures in the first session part; from the second session part onwards, two other patients begin to gesture and the overall number of gestures increases slightly over time (on average across all patients, the number of gestures amounts to (1) 0.5, (2) 0.5 and (3) 0.75 in the different session parts). Given that these numbers are still very low, it would remain to be seen in studies over several sessions, and with real patients, to what extent pointing gestures among patients increase over time, as reflecting processes of patient empowerment.

Furthermore, we hypothesized that within one session, there is a marked trend towards fluent documentation with TBM—as reflected by a decrease in data input problems (hypothesis H3.2).

Results: The number of recorded difficulties in the handling of TBM decreases over time. This is the case, even though the therapists dare to use more functionality over time—with new functionality occasionally bringing about new difficulties in the handling of the system. In session part one, an average of 5.5 entry difficulties are recorded per therapist. In session parts two and three, there are only 3 difficulties on the average.

Figure 9 shows a numerical representation of how therapists and patients interact with TBM across the four 50-min consultation sessions on the average. For the majority of time, therapists direct their full attention to the patient without interacting with TBM. The second biggest share of the therapist interaction is the input of information in the system. This comprises diverse input modes, such as digital whiteboard, keyboard and tablet computer. Also the patients direct their attention to the therapist for the biggest part of time. While they do not enter information in the TBM system themselves, they watch the TBM whiteboard interface actively and point at it occasionally.

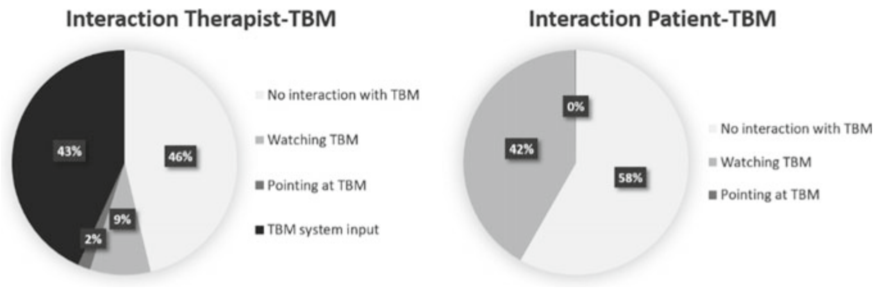


Fig. 9 Average distributions of the interaction modes with the Tele-Board MED (TBM) system in consultation sessions of 50 min length (N = 4). The numbers are rounded percentage values

4 Conclusion

This chapter describes experiences with the Tele-Board MED (TBM) system for collaborative note-taking in doctor-patient sessions in the domain of mental health-care. The goals of the TBM system used in medical consultations are: (1) supporting faster, high-quality documentation (2) supporting patient-doctor teamwork, and (3) providing a positive learning experience for users. We conducted a user experience study with behavior psychotherapists who used the TBM system in patient interview sessions. We assessed the therapists' usage of the TBM report generation feature, the therapists' experience when using TBM in a patient consultation, as well as the interaction dynamics between therapist, patient and the TBM system.

We found with the TBM medical report generation feature, therapists save 60% of the time they normally spend on writing reports to health insurance companies. This is even the case when the therapists used TBM for the first time with a patient. Thus, TBM unfolds its full potential after the treatment session by allowing the therapist to create official clinical documents automatically. We can conclude that TBM not only serves as an interface for information entry, but also facilitates therapeutic conversation, namely by serving as a common reference therapists and patients can watch and point at. Furthermore, we can conclude that even in the very first treatment session with TBM, therapists learn to integrate the system in the patient anamnesis interview smoothly. Observed problems of data entry due to software behavior decrease over time, which reflects a marked trend towards fluent documentation. The therapists' subjective feeling expressed by hand-drawn user experience curves also becomes clearly positive over time. The first third of the 50 min present a phase of the therapist's familiarization with the unconventional computer-supported consultation scenario. In fact, therapists not only need to become familiar with the system operation, but also with harmonizing the activities of open note-taking and therapeutic conversation. In the middle of the session, therapists encountered situations of uncertainty when it was unclear whether the information voiced by the patient was relevant to note down or when the expressed information was contradictory to their own perception. On the other hand, TBM creates an increased sense of security when therapists provide explanations to the patient, e.g. when introducing behavior analysis methods or when summarizing the session content. In the last third of the session, the therapists are feeling confident, calm and positive.

In summary, this user experience study shows that TBM supports various aspects of therapeutic conversations. During the therapeutic conversation, it serves as a tool to capture important information on the fly and to guide the therapist-patient dialogue. After the session, it serves as an assistant to create case reports automatically by summarizing the information captured in the session.

Acknowledgements We would like to express our gratefulness to the psychotherapists and the volunteer patient actors who participated in this study. We thank Dr. Sharon Nemeth for copyediting this chapter.

References

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development, 55*(4), 1278–1289.

Bundesgesetz. (2013). Gesetz zur Verbesserung der Rechte von Patientinnen und Patienten, Bonn.

Bundesvereinigung, K. (2017). Leitfaden zum Erstellen des Berichts an die Gutachterin oder den Gutachter, http://www.kbv.de/media/PTV3_Ausfuellhilfe.pdf.

Coyle, D., Doherty, G., Matthews, M., & Sharry, J. (2007). Computers in talk-based mental health interventions. *Interacting with Computers, 19*(4), 545–562.

Gottman, J. M., & Levenson, R. W. (2000). The timing of divorce: Predicting when a couple will divorce over a 14–year period. *Journal of Marriage and Family, 62*(3), 737–745.

Grawe, K. (2005). Empirisch validierte Wirkfaktoren statt Therapiemethoden. *Report Psychologie, 7*(8), 311.

Gumienny, R., Gericke, L., Quasthoff, M., Willems, C., & Meinel, C. (2011). Tele-Board: Enabling efficient collaboration in digital design spaces. In *Proceedings of the 15th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, (pp. 47–54).

Jung, M. F. (2011). *Engineering team performance and emotion: Affective interaction dynamics as indicators of design team performance*. Dissertation, Stanford University. Available online at: http://purl.stanford.edu/th996ft5752.

Kahn, M. W., Bell, S. K., Walker, J., & Delbanco, T. (2014). Let's show patients their mental health records. *JAMA, 311*(13), 1291–1292.

Knowles, S. E., Toms, G., Sanders, C., Bee, P., Lovell, K., Rennick-Egglestone, S., et al. (2014). Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. *PLoS ONE, 9*(1), 1–12.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011). UX curve: A method for evaluating long-term user experience. *Interacting with Computers, 23*(5), 473–483.

Peluso, P. R., Diaz, P., & Tabib, G. (2018). Quantitative video coding of therapist-client sessions. In U. Strawinska-Zanko & L. Liebovitch (Eds.), *Mathematical modeling of social relationships. Computational social sciences*. Cham: Springer.

Perlich, A., & Meinel, C. (2015). Automatic treatment session summaries in psychotherapy—A step towards therapist-patient cooperation. *Procedia Computer Science, 63,* 276–283.

Perlich, A., & Meinel, C. (2016). Patient-provider teamwork via cooperative note taking on Tele-Board MED. In *Exploring complexity in health: An interdisciplinary systems approach* (pp 117–121).

Perlich, A., & Meinel, C. (2018). Cooperative note-taking in psychotherapy sessions: An evaluation of the therapist's user experience with tele-board MED. In *Proceedings of the IEEE International Conference on e-Health Networking, Applications and Services (Healthcom)* (pp. 1–6).

Perlich, A., von Thienen, J. P. A., Wenzel, M., & Meinel, C. (2018). Learning from success and failure in healthcare innovation: The story of tele-board MED. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research: Making distinctions: collaboration versus cooperation* (pp. 327–345). Cham: Springer.

Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology, 79*(3), 284–295.

Reinecker, H. (1999). Fallbuch der klinischen Psychologie: Modelle psychischer Störungen, Einzelfallstudien zum Lehrbuch der klinischen Psychologie. In Hogrefe, Verl. für Psychologie, pp. 23–27, 73–78.

Topor, D. R., AhnAllen, C. G., Mulligan, E. A., & Dickey, C. C. (2017). Using video recordings of psychotherapy sessions in supervision: Strategies to reduce learner anxiety. *Academic Psychiatry: The Journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry, 41*(1), 40–43.

von Thienen, J. P. A., Perlich, A., & Meinel, C. (2015). Tele-Board MED: Supporting twenty-first century medicine for mutual benefit. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research: Building innovators* (pp. 101–130). Cham: Springer.

Wenzel, M., Perlich, A., von Thienen, J. P. A., & Meinel, C. (2019). New ways of data entry in doctor-patient encounters. In C. Meinel & L. Leifer (Eds.), *Design thinking research. Understanding innovation* (pp. 159–177). Cham: Springer.

Towards More Human-Centered openHPI Collab Spaces

Hanadi Traifeh, Thomas Staubitz and Christoph Meinel

Abstract Since their inception, Massive Open Online Courses (MOOCs) have generated significant attention from students and particularly lifelong learners. This has led to increased interest from academic institutions to develop their own online platforms or to collaborate with other existing platforms, such as edX, to offer their MOOCs. This chapter focuses on openHPI, the MOOC platform offered by the Hasso Plattner Institute in Germany. One of the platform's important features is the Collab Space, a virtual private space for groups and teams in which they can interact and collaborate on assignments and projects using a set of communication and collaboration tools. The conducted study examines the current state of the Collab Space from a learner's perspective by assessing the functionality of its communication and collaboration tools and how they are being used by the participants. We applied a design thinking approach to carry out the study and to develop solutions for some of the platform deficiencies revealed by the study. During the study, we observed teams while performing tasks and interacting together in the Collab Space, and evaluated how the teams used their tools. Semi-structured interviews were conducted during two stages of the study. We argue that by applying the design thinking methodology and putting participants at the center of our research, new insights on how to improve the user-centeredness of the Collab Space can be achieved. We conclude this chapter by outlining next steps for research and potential future opportunities.

H. Traifeh (✉) · T. Staubitz · C. Meinel
Hasso Plattner Institute for Digital Engineering, Prof.-Dr.-Helmert Straße 2-3, 14482 Potsdam, Germany
e-mail: hanadi.traifeh@hpi.de

T. Staubitz
e-mail: thomas.staubitz@hpi.de

C. Meinel
e-mail: christoph.meinel@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*,
Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_16

1 Introduction

openHPI is a MOOC platform offered by the Hasso Plattner Institute in Potsdam, Germany. The platform has been providing life-long learners with courses in a wider IT context since 2012 (Meinel and Willems 2013), and recently introduced its first MOOC on Design Thinking. openHPI offers courses in German and English. The basic structure of the openHPI courses follows the xMOOC model with structured learning activities such as video lectures, interactive self-tests and assignments. The platform, as well as most of the courses' teaching teams, encourage social interaction among students through the main discussion board of each course. Some of the teaching teams actively trigger the participants' large group collaboration in the course-wide discussion boards in various ways (Staubitz et al. 2018). Fewer courses further enrich this basic interaction model by emphasizing the social learning approach. Students are either encouraged to form interest groups of their own or, in some cases, the instructors form small teams in which the participants are asked to interact and collaborate while working on a given project. These interactions take place within the platform's Collab Space, a virtual private space for teams or groups equipped with a set of communication and collaboration tools.

As we will be differentiating between teams and groups throughout the rest of the chapter, we will shortly present our definition of these terms in the given context. Groups are a loosely coupled assembly of course participants that share a Collab Space based on a given commonality, e.g. speaking the same language (differing from the official course language), coming from the same school or company, or just knowing each other from a different context. Teams, in contrast, are a more tightly coupled assembly of course participants that share a Collab Space as they are collaboratively or cooperatively[1] working on a given task. While the members of groups, in the majority of cases, are either joining on their own or are invited by the participant who started the group, teams in most cases are formed by the instructors (Staubitz and Meinel 2017).

A study by Zheng et al. shows that although many MOOC platforms have tried to implement team-based learning, little collaborative success has been achieved (Zheng et al. 2015). Mak et al. also argue that despite the different learning activities within xMOOCs, these activities lack the beneficial group dynamics, especially if the students' interaction is limited to discussion forums (Mak et al. 2010). However, earlier research shows the advantages of group learning over individual learning on both cognitive and social levels (Baker and Lund 1997; Strijbos 2004). These advantages include increased attendance, improvement in academic results, and the development of social and team skills (Wen 2016). Other studies demonstrate that "deep learning and the development of critical and higher order thinking skills only occur through interaction and collaboration" (Staubitz et al. 2015; Brindley et al. 2009; Laal and Ghodsi 2012). Only a few of the current xMOOCs providers have implemented or are working on incorporating a collaborative team-based learning

[1] See (Staubitz and Meinel 2018a) for our definition of these terms.

component. One of the few examples besides openHPI, is NovoEd,[2] a platform that was established from the outset on a collaborative team- and project-based approach.

This chapter focuses on the openHPI MOOC platform in particular, and reports on our study in which we applied the design thinking methodology to evaluate the current state of the platform's Collab Space feature from a human-centred perspective. Our study examined how users work together within teams and to what extent the communication and collaboration tools provided in the Collab Space are serving the users' identified needs.

2 The Collab Space

The Collab Space feature (Fig. 1) of the openHPI MOOC platform was implemented in 2013 as one of the core features of the platform to offer student groups a private space in which they can interact with each other in a more private setting than the wilderness of the course forum (Staubitz et al. 2015). Later on, in 2016, the Collab Space was enhanced with the option of allowing instructors to add assessable team assignments to their courses. It is important to keep in mind here that both, the matching of the teams and the assessment of the teamwork need to be scalable as the courses on the openHPI platform often have tens of thousands of participants (Staubitz and Meinel 2017).

When we started our research, the following tools had been provided in the Collab Spaces:

- **Discussions**: A discussion forum where students can discuss topics, post questions and reply to those of their teammates. Other than the course-wide forum, this forum is only accessible for the members of the Collab Space and the instructors.
- **Etherpad**: A collaborative open source text editor similar to Google Docs.
- **Tele-Board**: An interactive virtual board where students can share ideas and do brainstorming.
- **Google Hangout**: Allows synchronous communication within the team.
- **Use openHPI Together**: Synchronizes the browsers of the participants in the session. Participants can also see each other's mouse movements.

At that time, the Collab Space interface was structured as follows. On the left side, a navigation bar includes from top to bottom: Dashboard, Files, Discussions, Peer Assessment, Etherpad, Tele-Board and Administration. The middle section is dedicated for viewing content-related to the option chosen from the navigation bar. The right side of this version of the Collab Space had two sections: A Hangout button for starting a video call, and another button for using openHPI together.

In some of openHPI's MOOCs, 'team work' is required to submit assignments and work on projects. Students are usually assigned to teams by the course administrators/teachers based on different criteria usually decided by the course instructor.

[2]https://novoed.com.

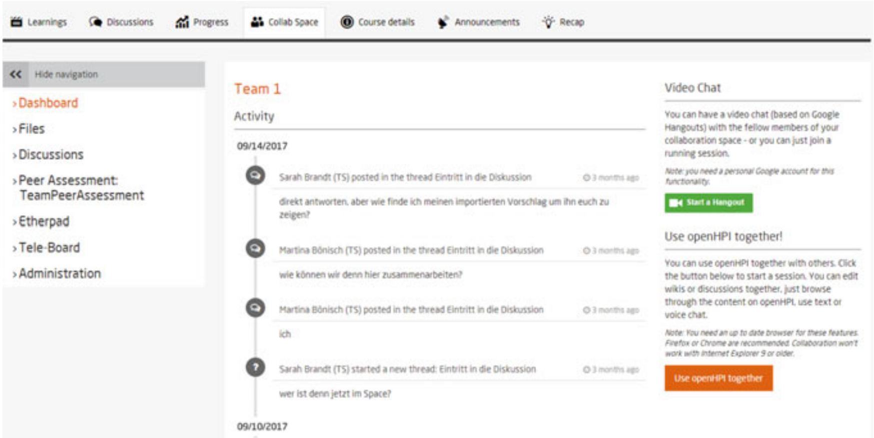


Fig. 1 The Collab Space interface

However, all of openHPI's courses—whether they require team work or not—offer students the option of forming learning groups themselves. These groups can be open or closed. Open groups can be joined by any student, while closed groups are controlled by their administrators and participants need to request membership. Both group types are accessible by the teaching teams and platform administrators.

3 Research Approach

There is little agreement in the literature about the exact definition of design thinking (DT) (Koh et al. 2015). Dunne and Martin suggest that "design thinking is the way designers think: the mental processes they use to design objects, services or systems, as distinct from the end results of elegant and useful products" (Dunne and Martin 2006). According to Brown, DT is "a human-centered approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success." (Brown 2009). Other researchers define design thinking as "a heuristic, a series of steps or as strategies that scaffold people to have the ability to solve complex or 'wicked' problems or to create an innovative product" (Koh et al. 2015; Razzouk and Shute 2012). MacFadyen suggests that "design thinking uses divergent and convergent thinking to 'flesh out' potential solutions for problems at any level" (Koh et al. 2015; MacFadyen 2014).

Design thinking has been widely implemented in different sectors such as Information Technology (IT), economics, education, government, healthcare, non-for-profit organizations and others. Many models have been developed using the methodology to address the challenges each sector faces. The essential first step in the design

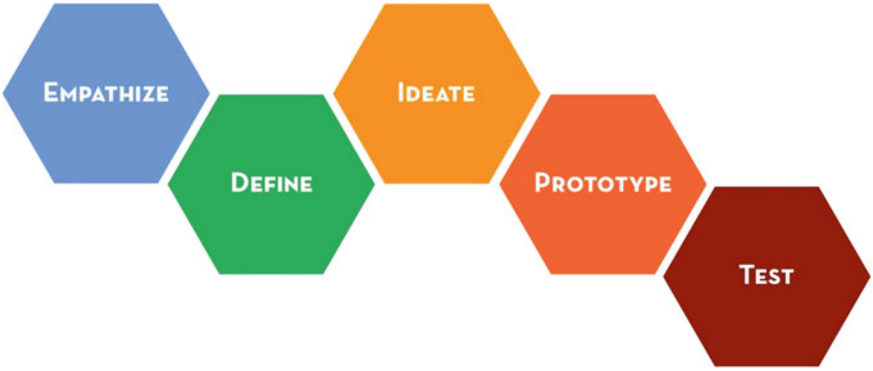


Fig. 2 The Design Thinking Model—Image by Stanford d-school (https://dschool.stanford.edu/resources/the-bootcamp-bootleg)

thinking process in all models focuses on *understanding the problem*. The process then proceeds to conceive and develop a solution that has the user's needs at its core.

For our research purposes, we will use a model developed by The Hasso Plattner Institute of Design at Stanford. According to Stanford's professor David Kelly, the process of design thinking consists of five steps: Empathize, Define, Ideate, Prototype and Test (Fig. 2). Although these five steps are often described in a linear way, design thinking is an iterative process (Mononen 2017) in which the designer can go back and forth to different phases throughout the process depending on the needs of the challenge that is being tackled.

In order to solve a problem, designers first need to understand the challenge and the users for whom they are designing. This understanding is built through *Empathy*. At the Empathy phase, which is "the centerpiece of a human-centered design process" ("An introduction to design thinking" 2010), designers observe the users within the context of the challenge to get a better understanding of their behavior, how they do things, or how they use the service/product designers intend to improve or redesign. Designers also interview users to understand their points of view, and may also immerse themselves in the challenge to appreciate the same experiences and feelings as those of users. The next step is to *Define* the users' needs and the challenge/problem itself. In the *Ideation* phase, designers challenge assumptions, and go wide in creating concepts and generating ideas for innovative solutions. This phase creates a smooth transition into the *Prototyping* phase in which designers create solutions based on the ideas that show potential, and *Test* them with the users to get direct feedback and see what worked well and what did not, so that they can iterate their solution until they reach a satisfactory outcome.

Since design thinking has been proven to be an effective approach to tackle challenges that involve human factors, we used it in our study to better understand the user experiences in openHPI's Collab Space and to redesign these experiences to better meet the users' real needs and expectations.

3.1 The MOOC 'Object-Oriented Programming in Java' as a Field of Experiment

Object-Oriented Programming in Java is a course offered in German on openHPI and ran from March 27, 2017 through May 14, 2017. Halfway through the course, 9242 participants were enrolled. By course end this number had risen to 10,402. The basic structure of the course consisted of instructional videos, each followed by a short multiple-choice test and three auto-graded programming assignments in several levels of difficulty. The tests were offered with the help of a browser-based auto-grading tool that has been developed by the openHPI team particularly for this purpose (Staubitz et al. 2016).

The teaching team strongly and successfully encouraged the participants to use the whole variety of the platform's communication tools to engage in social constructivist discussions to gain competence in programming. The success here can easily be measured in the comparably high forum participation and the high quality of discussion (Staubitz and Meinel 2018a).

Additionally, the course offered an optional team assignment on object-oriented modeling. This task offered only few bonus points and required a sufficiently high amount of work, which scared off many of the participants at the beginning. Despite this setting, about 1500 participants registered for the team assignment.

The assignment required team collaboration and the extensive use of the Collab Space by participants to work on their assignments, submit their final project and perform peer reviews at the end of the course. Teams in this course were formed in an interventionist way (Kizilcec 2013) by the instructors. Diversity of professional background, gender and age and homogeneity of the participants' time commitment for the given task served as the main matching criteria.

For our study, we wanted to assess the Collab Space feature from the user's perspective, by observing and asking a group of users how they use the space when collaborating with their teammates, which of the Collab Space tools they used the most and which they believe were not supporting their learning journey and team work. Our goal is to enable teams to collaborate more efficiently and to support them on their learning journey. We aim to improve the *user friendliness* and *user/human centeredness* of openHPI's Collab Space feature. For this purpose, we started by interviewing some of the participants who completed the course. We collected data from 14 participants, two of whom had belonged to the same team when they attended the course. Therefore, our data reflects the interaction among members of 13 teams in total. We interviewed 4 participants face to face, and the other 10 via video calls (6 via Skype, 4 via Hangout). We carried out semi-structured interviews with several open questions that focused on the participants' experience with the platform and with their teammates during the course. We also asked participants about each of the communication tools implemented in the Collab Space and how they used them.

Four of the interviewees are in their 40s, 4 are in their 50s, 3 are in their 30s and the last 3 are still in their teens (16, 17, 19 years old). While 6 of the participants come from an IT background, the others have the following backgrounds: high school, mechanical engineering, sales management, biology, media, and insurance. Most of the participants claim that programming is a hobby and that they joined the course to improve their skills or to learn more about java programming. This composition of the interviewees closely reflects the socio-demographic structure of the course participants in total. Immediately following each interview, and as part of the design thinking empathy phase, we ran an observation exercise. Participants were provided with usernames and passwords that were set up in advance by the researchers. Participants were instructed to log into the Collab Space and perform some tasks while the researchers were able to observe and record the participants' activities. The features of the Collab Space were explored and evaluated from the users' perspective and the results were recorded and analyzed. In total we spent a minimum of 1 h with each of the interviewees. The ratio between interview and observation within the total session differed from case to case.[3]

3.2 Feedback on the Different Collaboration Tools

Discussion Board: While the vast majority of participants (n = 11) used the discussion board, they stated that they used it either to arrange meetings or as a tool for communication (chat tool) because there are no other alternative tools for chatting. Participant said, "..we used the internal discussion space to say hello and chat regarding our assignments, ask questions and so on.." However, almost all participants mentioned that they were confused between the discussion board offered inside the Collab Space and that of the main course page (Fig. 3).
Etherpad: Many of the interviewed participants (n = 8) used the Etherpad feature and no major complaints were expressed (Fig. 4).
Hangout: Most of our participants (n = 11) did not use Hangout. While some claim to have privacy concerns, the majority used other video call tools from outside the platform, and insisted on the importance of having a video call option available for use. Few participants expressed their desire to use Hangout in particular but they did not know how to use it. One of the major obstacles is the need for a Google account to use the feature. Another issue that popped up regularly is the need to schedule a meeting before going to the hangout. Many of the interviewees stated that they just started a hangout and then wondered why nobody joined them there. They also expressed the desire to be able to see who else is currently online on openHPI. However, all of those who mentioned this issue (at some point during the interview or observation) realized that even if the option existed of seeing which other team member is currently online, they would need to schedule a meeting first before just going there.

[3]In addition to this user-centered qualitative approach, we have evaluated a quantitative survey among the teamwork participants, which has been published separately (Staubitz et al. 2018). Furthermore, the actual interaction of the participants as documented by the captured learning analytics data of our platform was evaluated and published separately (Staubitz and Meinel 2018b).

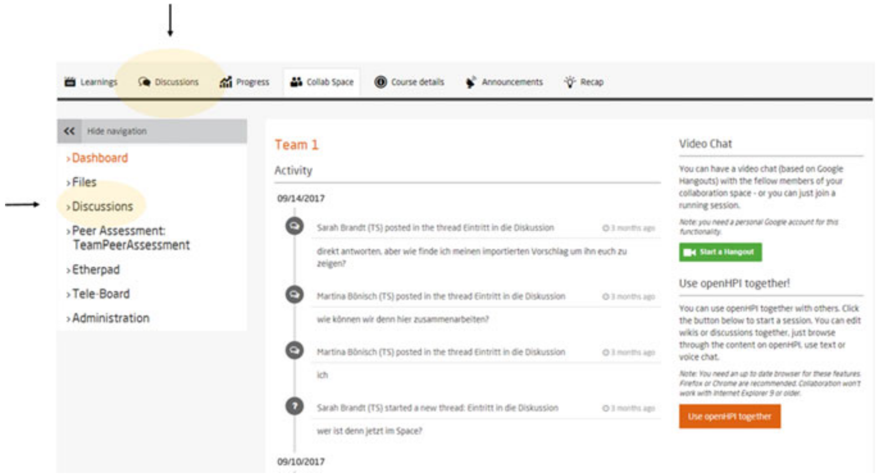


Fig. 3 Discussions board in the Collab Space

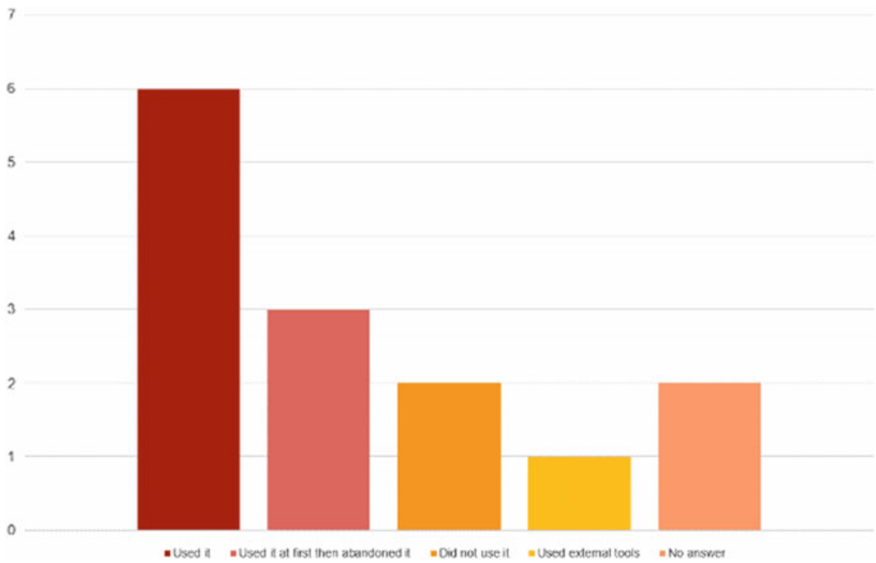


Fig. 4 Reported usage of the Etherpad feature

> We tried to use it (hangout) but it didn't work out because we couldn't agree on a time to talk.
>
> We could have used Google Hangouts, but we didn't really need it, and you also do not want to let everybody see your living room.

Tele-Board: Nine participants claimed that they tried to use the Tele-Board feature but found it too difficult or complicated to use. Two of those who used it said that it was not the optimal solution for course assignments.

> We (the team) used Tele-Board when we did our assignments. At first, I had to try it and had to learn how it works. And my first thought was it's too complicated!
>
> I didn't get what I can do with it!… Probably if we could use a pen instead of a mouse, it would have been easier.
>
> The Tele-Board is too complicated, I tried it but I never used it! … It would have been better if there had been another tool where we could use circles or lay out a model or that had the capability of drawing specific shapes.

We support the participants' view as the Tele-Board feature implemented on the platform is a web version which limits the feature's richness. Tele-Board is designed to be used on interactive boards and being operated with a computer mouse is not exactly its strength. The employed prototype also was not supported in use via smartphone or tablets, which would have offered a more natural way of interaction. Finally, the tool is not the optimal solution for all phases of the given task. While it has its benefits for the early stages of collaborative modeling, there are more suitable tools available to develop the final structure of the model.

Use openHPI together (Fig. 5): None of our interviewees used this feature. Even when they were asked to try it during our observation phase, participants had many problems and did not much like it.

3.3 Patterns and Findings

In design thinking, patterns from the gathered data are identified (Liedtka et al. 2013). Some hidden needs and real insights may be exposed by observing consistent and repeated expressions from different participants. By recognizing these patterns, potential solutions to the problem may start to emerge.

We identified *three* major categories based on the quotes made by interviewees and the recurrence of some direct requests or expression of needs.

1. **Team Interaction**: Many of our interviewees expressed that it was not clear how to start the initial conversation with their teams or how to work together. The following statements are a few examples:

   > Getting started was slow! Triggering the initial discussion would be helpful
   >
   > It was hard to find my team or to start communicating with them
   >
   > A little more info. on how to work together would be helpful!

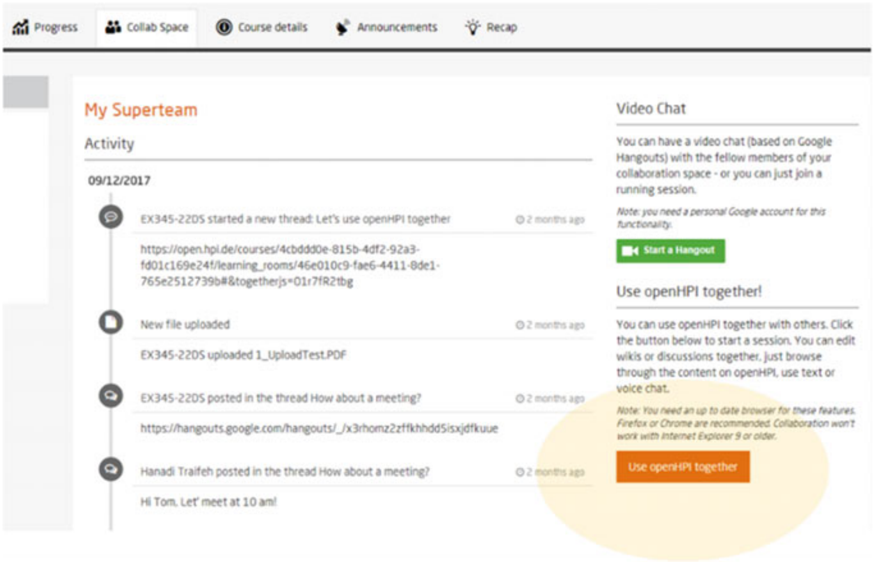


Fig. 5 Use openHPI together feature

2. **The Collab Space**: A large number of interviewees made clear statements about the design and structure of the Collab Space itself. They noted that entering the space was "overwhelming" and that they did not know which tool to use first or what exactly they were supposed to do once there.

   I looked through the platform and couldn't figure out what to do!

   One should have been exposed to the platform tools before starting the 'real' work!

   There was no explanation on "How to Use the Collab Space"

3. **Short introductory video**: The first two interviewees expressed their wish to have a short video on the platform that guides the participants through the use and functionalities of the different tools of the Collab Space.

   It is very important to know how to start using the Collab Space. Everybody should be able to understand the process. I wish there was a short video about this.

   It would be really helpful to have a video when you open the Collab Space that explains its features. This would definitely make life easier

We decided to act on this request and started to ask the other interviewees about the idea. The strong support for the idea expressed by the vast majority of the interviewees shifted our focus towards prototyping an introductory video and testing it as a next step (Fig. 6).

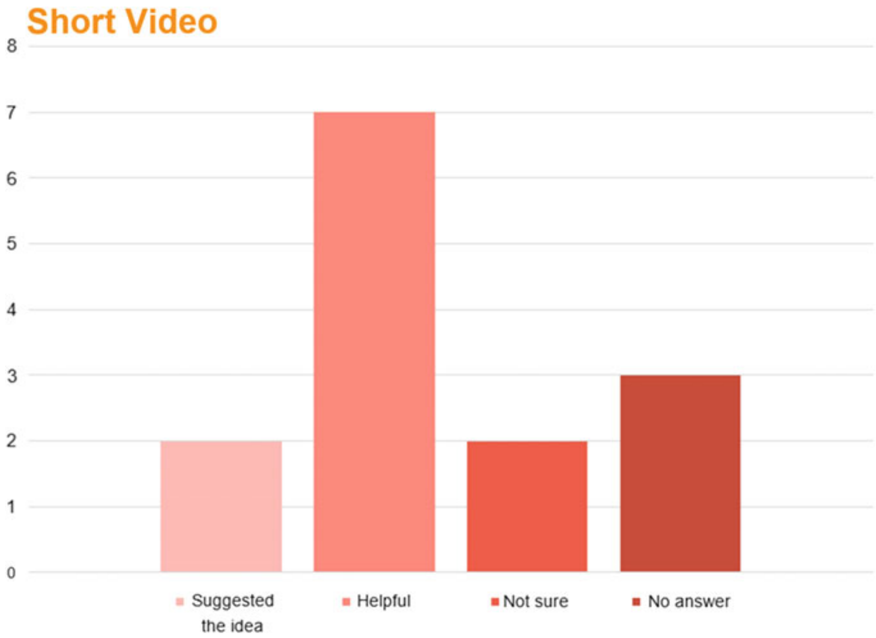


Fig. 6 Most interviewees agreed that the short introductory video would be helpful

3.4 Prototyping and Testing

We created a 7 min introductory video that introduces the features of the Collab Space. We then tested the video prototype in a workshop held during the d.confestival event, which took place at the Hasso Plattner Institute in Potsdam, Germany on September 14, 2017. We designed a new course titled 'Food is Life' that follows the same structure of openHPI MOOCs (Fig. 7) specifically for the purpose of the test and for conducting further research. We aimed for eight participants but only four joined the workshop. While this might be considered one of the limitations of our study, the small number of participants allowed for more time to discuss the collaborative team-based learning topic, do more observation and dive deeper into the users' needs. The participants were divided into two teams (2 members each) and were provided with laptops that had pre-set usernames and passwords to record their activities. None of the participants had attended any MOOC on the openHPI platform before, which makes this their first encounter with the platform. The new course (Food is Life) served as a base for the experiment. In the course, the participants' task was to collaborate in teams to create a dish recipe for a social gathering. One team was shown the video before performing the task. The other team logged into the platform and was expected to start working on the task right away.

Our observation shows that the team that watched the video performed better than the team that did not. For example, the team members who watched the video were

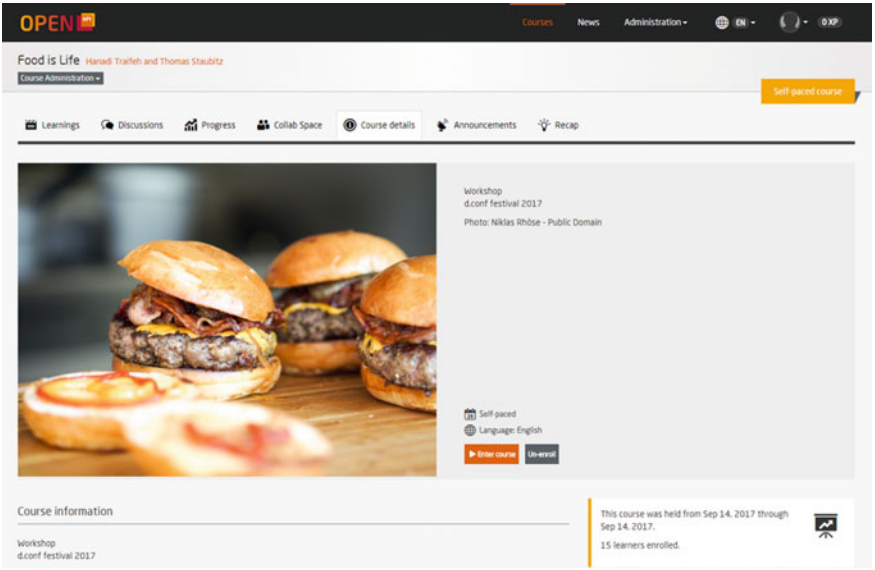


Fig. 7 ‘Food is Life’ course

able to find each other faster than those who did not. Moreover, the team that watched the video utilized the Etherpad feature right away, while it took the other team over 20 min to discover the tool. Overall, the ability to quickly use and benefit from the collaboration tools enabled the team that watched the video to complete the task (creating the dish recipe) successfully within the allotted time period (30 min). The team that did not watch the video clearly struggled in collaborating, and was finally not able to complete the task.

All participants also expressed their need for an instant messaging feature (a chat window). They believed that if that option were available, starting the initial conversation would have been much easier. Compared with the results of our first part of the study, we noticed that most of our participants used the ‘Discussion Board’ feature as a communication (chat) tool even though it did not have the full capability to serve this purpose. This highlights the importance of either including a chat tool within the platform or enhancing the discussion board with more chat-like features to satisfy users’ needs.

4 The Iteration Phase

Based on the results of our study, we took further steps to improve the experience of using the Collab Space for openHPI users:

1. Removal of the ‘Use openHPI together’ section and feature permanently as none of the first or second groups of users have ever used it or reported any willingness to do so. Observing the participants struggling with the feature and our own difficulties using the feature made the decision to remove it relatively easy.
2. The Hangout feature is moved to the navigation bar on the left side and has been re-named “Start a Video Chat”. An additional page explaining the need to schedule a meeting before starting the hangout video call and providing some technical instructions on how to sign into hangout was added as part of this change. The original ‘Video Chat’ section and the ‘Start a Hangout’ button have been removed.
3. An enhanced version of the Collab Space introductory video was recorded to be introduced at the beginning of all courses. The initial version that was used in our experiment was only a prototype to validate our assumptions about the importance of including an introductory video. Participants also gave valuable feedback about the video itself, which was incorporated in our next iteration.
4. Recently, a commercial version of the Tele-Board was developed. While the Tele-Board has been a research prototype, the NexBoard builds on the results of this research and adds the advantages of a commercially offered software, such as maintenance, support and improved stability. One of the most important improvements that comes with this switch to the new tool is improved support for tablets. The course administrator always had the option of turning the feature on or off according to the course content. In the future, there will also be improved administration options, such as the provision of different templates for different courses.
5. A new wording structure replaced the old version of the navigation bar items. Verbs are used instead of abstract words that confused the users. The sentences themselves were also simplified. Table 1 gives a few examples.
6. The items of the navigation menu have been rearranged according to their logical order (Fig. 8).

Table 1 Old and new structure of the navigation bar

Old	New
Dashboard	View recent activity
Discussions	Discuss with your team
Etherpad	Collaborate on texts
Tele-Board	Brainstorm ideas
(Hangout)	Start a video chat
Files	Share Files
Peer Assessment: [Name of the assignment]	Submit your team work and evaluate your peers
Administration	Manage your Collab Space

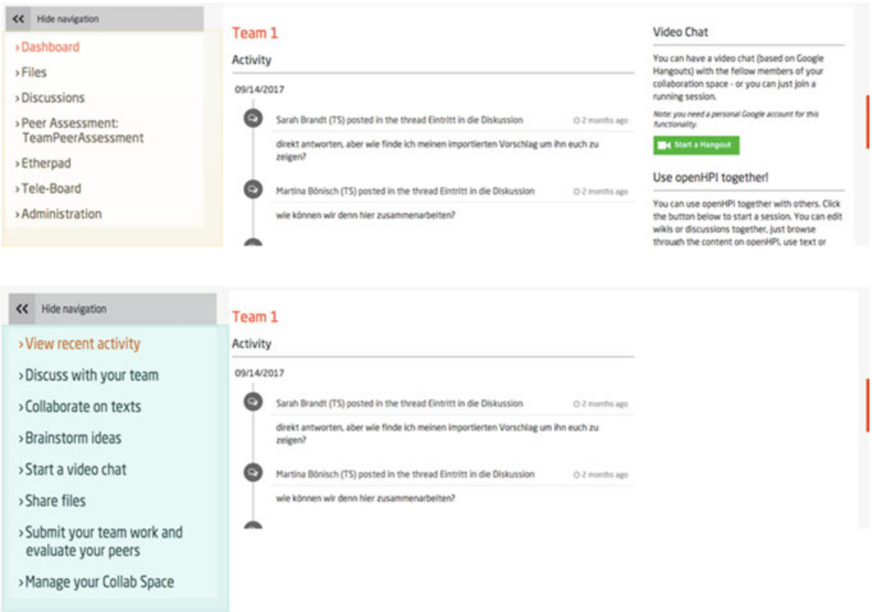


Fig. 8 Re-wording and rearranging of the Collab Space items

7. Although having an 'instant messaging/chat' feature seemed important to our users, we decided to develop and probably implement this feature in a next iteration.

The new design of the Collab Space has been tested with new teams through a MOOC called Intrapreneurship—Make your business great again :-)[4] (30 October 2017–26 March 2018). This course featured two tracks: a fast track and a full track. The full track added a small team-based project on top of the fast track. The participants had to develop a pitch for fictitious business ideas that were provided by the course participants themselves. Working on the pitch required the usage of the Collab Space. Furthermore, we are working with the new Collab Space in an iteration of the Java course, which is offered in a version that has been particularly refined to serve the need of schools: Object-oriented programming in Java—School-Cloud-Edition 2018[5] (26 February 2018–11 June 2018)

The results of these tests are still being evaluated and will be published separately.

Once these courses have been evaluated, further iteration may be carried out after gathering new feedback and observing the new users' experience within the platform.

[4] https://mooc.house/courses/bizmooc2018.

[5] https://open.hpi.de/courses/javaeinstieg-mint-ec-2018.

5 Conclusion

This chapter presented an assessment of the current state of the Collab Space at openHPI MOOC platform from a human-centered design perspective. A design thinking (DT) approach was followed to evaluate the functionality of the different communication tools implemented in the Collab Space. Our study results show that DT facilitated the discovery of the Collab Space users' needs and the development of better solutions that may encourage teams to work better together and improve their learning experience. The prototype designs we tested following the DT approach resulted in clear improvement in user collaboration and better engagement with the openHPI MOOC platform.

We aim to continue our exploration and the prototyping of iterative designs for improving the engagement models and approaches on MOOC platforms in particular, and in digital learning in general. We will support our DT-based study with quantitative evidence from the platform's usage statics and logs. We also intend to expand the scope of our studies to further validate our findings and test our insights through a more diverse participant population.

References

An Introduction to Design Thinking PROCESS GUIDE. (2010). The Hasso Plattner Institute of Design at Stanford. Retrieved from https://dschool-old.stanford.edu/sandbox/groups/designresources/wiki/36873/attachments/74b3d/ModeGuideBOOTCAMP2010L.pdf.

Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of Computer Assisted Learning, 13*(3), 175–193.

Brindley, J., Blaschke, L. M., & Walti, C. (2009). Creating effective collaborative learning groups in an online environment. *The International Review of Research in Open and Distributed Learning*, *10*(3).

Brown, T. (2009). *Change by design: How design thinking creates new alternatives for business and society*. Collins Business.

Dunne, D., & Martin, R. (2006). Design thinking and how it will change management education: An interview and discussion. *Academy of Management Learning & Education, 5*(4), 512–523.

Kizilcec, R. F. (2013, June). Collaborative learning in geographically distributed and in-person groups. In *AIED 2013 Workshops Proceedings Volume* (Vol. 67).

Koh, J. H. L., Chai, C. S., Wong, B., & Hong, H. Y. (2015). Design thinking and education. In *Design thinking for education* (pp. 1–15). Springer, Singapore.

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia-Social and Behavioral Sciences, 31,* 486–490.

Liedtka, J., King, A., & Bennett, K. (2013). *Solving problems with design thinking: Ten stories of what works*. Columbia University Press.

MacFadyen, J. S. (2014). Design thinking. *Holistic Nursing Practice, 28*(1), 3–5.

Mak, S., Williams, R., & Mackness, J. (2010). Blogs and forums as communication and learning tools in a MOOC. In *Proceedings of the 7th International Conference on Networked Learning.* University of Lancaster.

Meinel, C., & Willems, C. (2013). *openHPI: das MOOC-Angebot des Hasso-Plattner-Instituts*. Technical report.

Mononen, L. (2017). Systems thinking and its contribution to understanding future designer thinking. *The Design Journal, 20*(sup1), S4529–S4538.
Razzouk, R., & Shute, V. (2012). What is design thinking and why is it important? *Review of Educational Research, 82*(3), 330–348.
Staubitz, T., Pfeiffer, T., Renz, J., Willems, C., & Meinel, C. (2015). Collaborative learning in a MOOC environment. In *Proceedings of the 8th Annual International Conference of Education, Research and Innovation* (pp. 8237–8246).
Staubitz, T., Klement, H., Teusner, R., Renz, J., & Meinel, C. (2016). CodeOcean—A versatile platform for practical programming exercises in online environments. In *Proceedings of the 2016 IEEE Global Engineering Education Conference (EDUCON)* (pp. 314–323).
Staubitz, T., & Meinel, C. (2017). Collaboration and teamwork on a MOOC platform: A toolset. In *Proceedings of the Fourth (2017) ACM Conference on Learning@Scale* (pp. 165–168).
Staubitz, T., & Meinel, C. (2018a). Collaborative learning in MOOC—Approaches and experiments. In *Proceedings of the 2018 IEEE Frontiers in Education Conference (FIE).*
Staubitz, T., & Meinel, C. (2018b). Team Based assignments in MOOCs: Results and observations. In *Proceedings of the Fifth (2018) ACM Conference on Learning@Scale* (pp. 47:1–47:4).
Staubitz, T., Traifeh, H., & Meinel, C. (2018). Team-based assignments in MOOCs—User feedback. In *Proceedings of the 2018 Learning with MOOCS (LWMOOCS)* (pp. 39–42).
Strijbos, J.W. (2004). The effect of roles on computer-supported collaborative learning.
Wen, M. (2016). *Investigating virtual teams in massive open online courses*: *Deliberation-based virtual team formation, discussion mining and support* (Doctoral dissertation, Ph.D. thesis, CMU).
Zheng, Z., Vogelsang, T., & Pinkwart, N. (2015). The impact of small learning group composition on student engagement and success in a MOOC. In *Proceedings of the 8th International Conference of Educational Data Mining* (pp. 500–503).

Overcoming Prominent Pitfalls of Work Space (Re-)Design: Using a Theoretical Perspective to Reflect and Shape Practice

Martin Schwemmle, Marie Klooker, Claudia Nicolai and Ulrich Weinberg

Abstract Companies are recognizing more and more the potential of work spaces to increase their employees' creativity, well-being, and performance. However, (re-)designing work space is not an easy task and, hence, many companies ask for recommendations and advice from a researcher's perspective. Taking these rising demands and experiences with work space design projects in organizations as a starting point, we link a practitioner's with a researcher's view. In this chapter we present prominent pitfalls of workspace (re-)design through the lens of theories, selected concepts, and frameworks. In particular, we deal with the three issues of (1) a too narrow understanding of work space, (2) a lack of understanding of the status quo, and (3) a missing awareness of the behavioral component of space and its potential for change management. Perspectives from theory give enhanced reflection on each issue. In addition, tools and examples allow us to transfer the theoretical knowledge into action in form of workshops or initiatives as part of a space (re-)design project. Researchers receive an overview of the issues that are relevant for practitioners and which also indicate topics for further research. Practitioners are provided with a better understanding of relevant concepts and actionable tools for their work space-related projects.

1 Introduction

Numerous companies loose potential with regard to their employees' creativity, well-being, and performance by not leveraging the work space as an organizational resource (Kristensen 2004; Elsbach and Pratt 2007). Recent studies show that many employees encounter issues with their work environment and feel dissatisfied with it with regard to its effects on productivity and well-being and its support for their work-related tasks (Gensler 2016; Steelcase 2016). However, companies are beginning to recognize the untapped potential of the work place and are starting to build

M. Schwemmle (✉) · M. Klooker · C. Nicolai · U. Weinberg
HPI School of Design Thinking, Hasso Plattner Institute for Digital Engineering, Campus Griebnitzsee, 14482 Potsdam, Germany
e-mail: martin.schwemmle@hpi.de

C. Meinel and L. Leifer (eds.), *Design Thinking Research*, Understanding Innovation, https://doi.org/10.1007/978-3-030-28960-7_17

new campuses or rebuild their existing offices, accordingly (e.g., Laing et al. 2011; Bacevice et al. 2016). Our own experience at the HPI School of Design Thinking also shows that the number of both, companies and non-profit organizations, who seek our support in regard to (work) space-related challenges, has been growing progressively during the last years. We use these encounters with practitioners, and the topics and questions guiding our talks, discussions, and workshops with them, as a starting point for this book chapter. Informed by our practical experience, we seek to bridge the gap between what "actually happens" in many space (re-) design projects and the frameworks and perspectives research offers to reflect these phenomena (Orlikowski 2015, p. 33). Linking a practitioner's with a researcher's view allows for a better understanding and also improvement of organizational reality.

According to Schön's Reflective Practitioner, "[i]n order to solve a problem by the application of existing theory …, a practitioner must be able to map those categories onto features of the practice situation" (Schön 1983, p. 41). Thus, we have identified three interrelated issues practitioners encounter when redesigning work spaces. We elaborate on these three issues in more detail and, in the following sections, we will discuss them along theoretical frameworks and research findings. Lastly, we end this chapter with a discussion.

The three issues we have identified are (1) a limited perspective on work space design in general, (2) the lack of understanding the status quo (i.e., the current space use), and (3) a missing awareness of the behavioral component of space and its potential for change management. First, companies take a too narrow perspective on space. More precisely, they do not focus on the space itself—the term *space* originates from the Latin word *spatium*, which translates as room, area, or distance—instead they focus on its borders, such as walls and separations. As a consequence, they consider the work space only as an architectural topic of floorplans and square meters—a perspective also called "container theory of space" (Hofer 1996). While these components are crucial—there is no physical room without walls—a broader focus is required, in particular an understanding for the people using the space and their behaviors and emotional reactions in that space. To this end, we will introduce Lefebvre's interactionist understanding of space, the concepts of psychological ownership and affordances, and further provide an example for redesigning a very common everyday space—the kitchen.

Second, and linked with the first issue, many organizational space projects lack an understanding of the status quo. Instead of analyzing current space usage and usage patterns, many planners ask for simple how-to-guidelines with square meters per person or even for order lists with furniture. We argue that such top-down space concepts possibly fail or remain below tapping into the potential of leveraging productivity and well-being, since they ignore important context factors, such as the users, their behavior, as well as the corporate environment they are acting in. In contrast, we suggest a more bottom-up approach that starts with examining how work spaces are currently used and aims at developing a spatial strategy out of actual behavior—a perspective that is known as 'practice turn' in the strategy literature (Whittington 2006). We will propose several tools from the fields of management, marketing, and design and show how to apply them for questions of the work space.

Third, and as a consequence of the two before-mentioned shortcomings, many space projects underestimate the behavioral and cultural aspect of space and, thus, run too short when it comes to creating a substantial impact. For most projects, the opening of a new space is the end of the space (re-)building initiative. We consider such an early end as problematic and propose two stages: first, the physical building of the space and second, the behavioral change connected to it. The second stage includes the cultural change as well as establishing and learning new work (place) behaviors. It can be summarized as the change management process connected to the physical space building. Without changing peoples' minds and rituals of usage, a new space can never unfold its full potential. At the same time, a space can already be changed even without buying new furniture (e.g., by hacking existing spaces). We will introduce several examples of such behavioral changes on different levels that aim at changing a space.

Overall, this book chapter offers valuable insights for both researchers and practitioners. Researchers will find an overview of common misunderstandings and shortcomings practitioners face when approaching the topic of work spaces and, hence, get inspirations for further research in this field. They will learn about research-based frameworks and theories, which have been helpful to structure and understand important issues in the field of work spaces. In addition, the reflections and suggestions in this chapter can be considered as interventions that change the thinking and behavior of people in an organization and they offer also additional starting points for further research. Practitioners receive helpful new perspectives that challenge common assumptions about the topic of work space. In particular, we provide inspiring examples and frameworks that help both managers and project leaders to avoid mistakes when it comes to (re-)designing work spaces. Additionally, this chapter offers useful arguments to underline the relevance of the topic in front of supervisors and team colleagues and to avoid common pitfalls. Lastly, this chapter provides them with both the theoretical foundations of tools, theories, and terminology in the field of work spaces, and with explanations and examples of how to best apply this theoretical knowledge in practice. Equipped with this knowledge, we hope that the following pages advance the academic discussion about work spaces and help practitioners to better leverage the power of the work space for their organization and employees.

2 Perspectives on Space

In this section, we will introduce four concepts that broaden the understanding of space beyond walls, partitions, and square meters: Lefebvre's understanding of space, the concepts of ownership and of affordances, as well as the case of the Frankfurt kitchen.

2.1 *Interactionist Understanding of Space*

Approaching the question of what space is, we suggest the following framework of space, which relies on the work of Henri Lefebvre and Martina Löw and is shown in Fig. 1 (Lefebvre 1991; Löw 2008; Schwemmle et al. 2018). Following this understanding, a space is not only characterized through its spatial elements—buildings, rooms, walls, partitions—but through the ways that users interact with the space. It is only the interaction of these users with the spatial elements that create the space.

The understanding of space in terms of the users and their interaction with spatial elements broadens the concept of space by two important aspects. First, walls only have a height, width, and material qualities (perceived constructed space), but users as human beings have emotions and shared experiences with and within a space (reflexive construed space). Thus, while a sole quantitatively-oriented understanding of space relies on distances, proximities, and measurements, a more user-centered perspective asks which emotional states and behaviors these measured quantities trigger, how the atmosphere of the space impacts the users, and how the space hosts certain activities and allows common experiences. For instance, a pure numerical perspective might just imply that the larger an office is the better it is for the employee. While, on the one hand, a large office communicates status and might be just more convenient, on the other hand, it might lead to the user feeling lost or overwhelmed with empty space that they fill with visual clutter. Further, a small room does not necessarily cause negative emotions: Many students live in shared apartments with relatively small rooms. When asked about their student life, many people did not remember the small size of the room, but of the experiences they had there: focused-work on their thesis, cool talks with friends, parties with 20 people in the small kitchen, celebrating the submission of a thesis, etc. In a similar vein, work spaces are

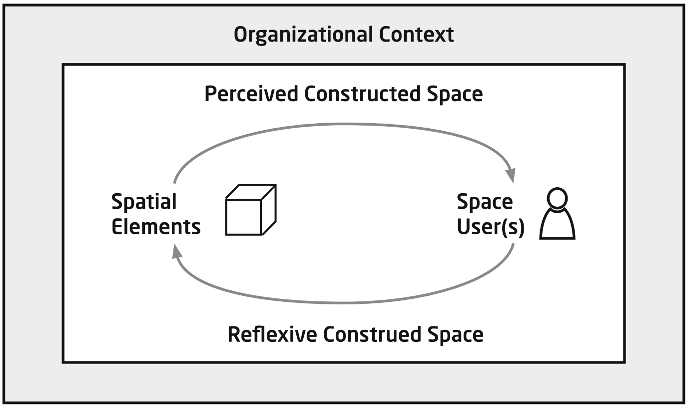


Fig. 1 The interaction of spatial elements with users compose the space (based on Lefebvre 1991; Löw 2008)

also characterized by the emotions of their users and the interactions with colleagues or customers in this space.

Second, a space that comprises both physical aspects and people offers two options for change. Changing a space can first be achieved by changing the physical components, which is what most people have in mind. However, changing a space can also mean that the behavior of people and the way they interact with the space is altered, for instance through spatial interventions (Klooker et al. 2016). To sum it up, this broadened and dynamic perspective of a space with spatial elements and users provides the foundation for a much broader understanding of the topic of space and also underlies many other concepts introduced in this chapter.

2.2 Psychological Ownership

We would like to start with a story to introduce psychological ownership. Some time ago, we ran a workshop supporting a company in developing new office spaces for their employees. On the first workshop day, we conducted interviews with the future users of the new offices and were welcomed with this statement: "So, you are the guys who want to take away my desk?" This quote illustrates two things. First, many employees consider work space design as an equivalent for abolishing personal desks and introducing so called hot desking. And second, that this person referred to the desk she was working at as *her* desk, while it was quite obvious that it had been paid for and was owned by the company she was working for.

The phenomenon of feeling as the owner of something without being its actual legal owner is called psychological ownership (Pierce et al. 2011). Philosopher Jean-Paul Sartre differentiates three categories of human existence—'to do', 'to be', and: 'to have' (Sartre 1956). Thus, the feeling of possessing something, the 'to have', is core for human beings. Research suggests that psychological ownership serves three fundamental human needs—self-efficacy, self-identity, and belongingness (Dawkins et al. 2017). All these three human needs also apply to the psychological ownership of the work space by its user. Self-efficacy refers to the need to feel capable in a given domain (Bandura 1997) and the psychological ownership of a work space including its furniture and privacy give space users the feeling that they can do the job. The two aspects therefore include having the documents, materials, and tools at hand, as well as creating the atmosphere required to be productive. Self-identity, our understanding of who we are, is also strengthened by psychological ownership. If 'my' desk is clean and organized or 'my' office has a friendly atmosphere through personal items, at the same time, this desk or office is communicating who I am as the psychological owner. This communication works in both directions—to myself and to others. Lastly, psychological ownership of a work space allows belongingness and it provides a home at work, as this quote by a manager participating in the same workshop nicely illustrates: "Basically, I don't need my own desk—I am in meetings the whole day, anyway. But there is this one thing: Every morning, I place my bag next to my desk and return to it in between meetings as my 'home base'."

Returning to the quote at the beginning of this paragraph, hot desking makes sense from an architectural or space usage perspective. However, from an ownership perspective, it takes away an employee's home. Does this imply that everybody needs his or her own desk all the time? Surely not. But it highlights the necessity to actively address and consider these issues and provide alternatives. For instance, in many companies with hot desking, employees personalize their lockers or have trolleys with some personal things as an alternative to having their own desk.

2.3 Affordances

One of the so-called ten principles of good design by designer Dieter Rams is that "Good design makes a product understandable" (de Jong 2017). This means, for example, in using a new smartphone you would ideally understand which gestures to apply, and which buttons to push without reading a manual. In a similar vein, when entering a new building as a visitor, there should be an indication of where to go, but also where it is not allowed to go. And, ideally, a new office is also self-explanatory with regard to personal and communal space, work and breakout zones, and focus and meeting areas.

These characteristics of a product or space explaining its own usage by design features are captured by the concept of affordances. According to literature, an affordance is defined as "what [the environment] offers the animal, what it provides or furnishes" (Gibson 1979, p. 127). Then, "[t]he animal (or its brain) performs inferences on the sensation, yielding a meaningful perception" (Chemero 2003, p. 181). Coming back to product design, examples for affordances are buttons that are designed in a way that it becomes clear that they have to be pushed, or knobs that communicate that they have to be turned (Norman 1988). In an analogy, a work space can also offer affordances to its users and thereby influence their behavior. Linking back to the interactionist understanding of space that we have introduced earlier, affordances highlight the fact that not only a space user is interacting with a space, but also, that a space is interacting with its users and communicating with them.

An example for using affordances to understand work spaces is to consider the informal interactions afforded by a place. Referring to a study by Fayard and Weeks (2007), the room where the photocopier is located in an office building can be an affordance for informal interactions. To fulfill this function, it should be large enough to host several people, be enclosed enough to mark the difference between outside and inside, and people should be comfortable to spend time there. In a similar vein, the welcoming design of social zones can communicate (afford) to space users that they are invited to stay there longer with others. Further, a large window or glass wall in a creative space is an affordance for people passing by to look inside and see what's going on there.

Affordances are also helpful to explain why creative spaces or innovation labs often have a design that is different from other rooms and work zones of the company. Such an extremely different design can be an affordance to change thinking and behavior or, in other words, a trigger to behave differently. In particular, garage-like spaces or maker spaces with work benches are an affordance to get active, to build and tinker.

In many cases, affordances work subconsciously—a comfortable-looking chair in a welcoming environment does not need a sign to invite somebody to sit on. Thus, it is important to both actively create affordances when redesigning spaces, but also to be aware that a lot of features might turn into unintended affordances. Coming back to the definition of affordances and the interactionist understanding of space, affordances are somehow in between objects and people. While the object offers something, peoples' brains have to make meaning out of it. Thus, different people—for instance employees versus visitors, long-time employees versus new hires—might draw different conclusions from the same affordance. It is therefore important to keep in mind the different space users and their backgrounds when considering affordances in a space.

2.4 The Frankfurt Kitchen: Why Size Does Not Always Matter

Many work space projects begin with questions about square meters per person or, even worse, start with the notion that "we don't have enough space for such fancy things as community zones or focus rooms." Such a purely quantitative perspective on the work space, however, runs too short, as the following historical development shows.

After World War I, there was a shortage of apartments for the German population, particularly in large cities. To mitigate this shortage, the German authorities wanted to build new affordable rental apartments for the working class. This requirement urged urban planners and architects to use space very consciously during the planning process. A very prominent example is how Austrian architect Margarete Schütte-Lihotzky redesigned the kitchen for a housing project in Frankfurt to use its space most efficiently. Following a movement that applied 'scientific management' to the household, a kitchen should be considered as a place for the preparation for food and "all unrelated work … should be kept out of the kitchen as much as possible." In essence, "a kitchen … can be much smaller than was formerly the case when it was used as a combined sitting-room, laundry and general workshop" (Frederick 1915, p. 19). Inspired by this notion and after analyzing kitchens in trains, Schütte-Lihotzky developed her so-called Frankfurt kitchen. She optimized the arrangement of objects in the kitchen based on the minimal necessary processes of food and meal preparation and thereby almost halved the kitchen size from 12.6 to 6.5 m^2 (please refer to Fig. 2). The Frankfurt kitchen became a huge success and was built over 10,000 times. It has further been exhibited in many design museums, including the MoMA (Museum of Modern Art 2010).

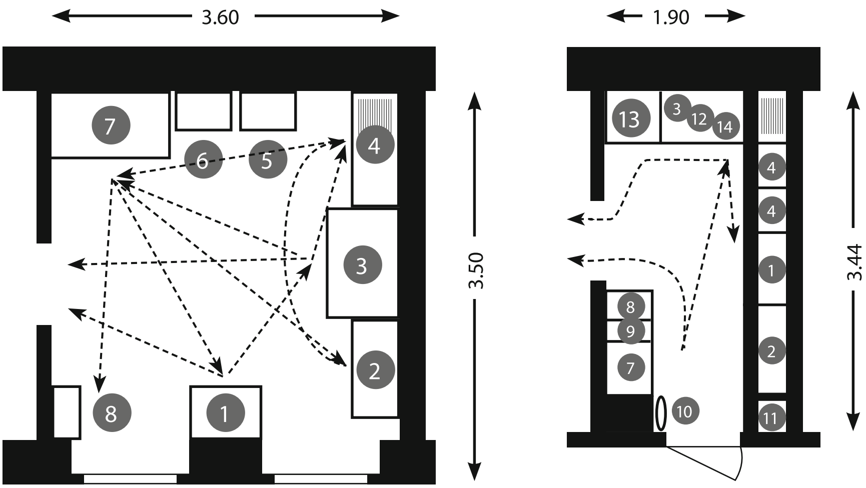


Fig. 2 Traditional (left) and Frankfurt kitchen (right) (adapted from Noever 1992, p. 11). 1: Aluminium storage bins, 2: cupboard for pots and pans, 3: work surface, 4: sink, 5: wooden box, 6: cupboard, 7: gas stove, 8: sideboard, 9: countertop, 10: heater, 11: broom closet, 12: swivel stool, 13: food cupboard, 14: garbage drawer

How can this almost 100-year-old story help us in designing work spaces today? We want to highlight two important learnings from the example of the Frankfurt kitchen. On the one hand, it is important to start by analyzing and understanding a space's function and the activities it should host before talking about size and designing the space. Thus, to begin a space redesign project, carefully observe the current usage with all relevant user groups—including, for instance, customers, guests, and facility management. Then, together with the users, analyze and discuss how these usage patterns reinforce each other, overlap, interfere, contradict, and therefore can be optimized, and, as a last step, define how these processes can be transformed into a space.

On the other hand, make conscious decisions. As we have shown in the example above, Schütte-Lihotzky made the conscious decision to reduce the kitchen to a place for food preparation. This separation of "the social space of the family and the workplace of women" (Jerram 2011, p. 541) was against the traditional German working-class style of building apartments. However, such a conscious decision was the basis for this new concept of work. If you introduce flexible zones in your new work environment, a desk in a focus zone might only be used for concentrated work and no longer be the place where any meeting or even short conversation takes place. Hence, the introduction of such new zones in an office requires a decision first (e.g., no meetings at desks) and, later on, a consequent cultural change (e.g., moving to a meeting zone instead of staying at the desk for conversations). Otherwise, the concept does not work. Accordingly, squeezing social life back into the tiny Frankfurt kitchen would not work.

Our experience fully confirms the effects of proper planning on a much better usage of space. To give an example, in a workshop we started with an existing floor plan that had been based on an average office size per worker. As a result of the workshop, built in the same size area there were work places for almost double the amount of people, although we had not focused, at all on optimizing square meter-usage per employee during the workshop. Instead, we started with the current usage of space and the processes and patterns that worked and did not work and then built the space around these needs. We will discuss this component of space, meaning the proper analysis of the status quo, in the next section.

As a last remark about the Frankfurt kitchen, this example should be carefully reflected in the context of ways of working. One aspect of this 'new work' movement is that work is no longer tied to a desk and working not only takes place in a dedicated room called an office. In this regard, tying an activity to a single room (work takes place at a desk in the office) is not the morale of this example. However, the Frankfurt kitchen is totally in line with 'new work' if considered in a broader context. Until its inception, the kitchen was considered a sitting-room, laundry, general workshop, and place to prepare food. Carefully analyzing these usage patterns and allocating these activities to different zones (the kitchen is for cooking only), fits to approaches of new work where the different activities of work take place in adequate environments. For instance, a desk is only used for concentrated work alone, and collaborative work takes place in places where collaboration is possible. At best, these different modes can be changed easily and, for example, a single desk is flexible and movable enough to become part of a team table.

The kitchen example highlights the relevance of analyzing a space usage first to then derive the spatial concept based on this analysis and how to make conscious decisions about minimal requirements. The next section will now focus on the topic of analyzing the status quo.

3 Analyzing the Status Quo

Having broadened the general perspective on work spaces in the last chapter, we now focus on the first step of the space (re-)design process—a proper assessment of the status quo, i.e., the current space usage. We therefore introduce the jobs to be done-framework, user journeys, and theory-grounded, and hands-on exercises to externalize latent needs.

3.1 Jobs to Be Done: Work Space Beyond Functionality

What is the function of a desk? A lot of people might say, it is a table to sit at and to write on; the Oxford Dictionary tells us something similar. Now have a look at your desk and think of what you do there. How often do you really write on it? And do you really need a table to put your laptop computer on? And, what about the pictures of your kids on the desk? And the snowball you won at last year's lottery?

Like a room is much more than walls, a desk is also much more than a table for writing. A framework focusing on exactly this variety of functions calls them "jobs to be done" and its inventor Clay Christensen gives the following example in his book (Christensen et al. 2016). A company specializing in apartments for downsizers—i.e., people who were seeking smaller apartments because their children had moved out—was facing a problem. Their beautiful new apartments could not find buyers. Christensen engaged in user research and discovered that the dining rooms of these apartments were far too small for the old dining room tables the downsizers brought with them. But why would somebody who moves out of their old family home bring the old huge family dining room table with them into a new apartment?

This question targets the issue of the job to be done. What is the job of a dining table? Certainly, to set with plates on and allow people to eat. Christensen calls this a functional job. However, this functional job could also be done by a new, much smaller table that would easily fit into the new apartment. But, this table has two more jobs—which Christensen calls an emotional and a social job. The emotional job of the table is connected to the many Thanksgiving and Christmas dinners, when the (large) family seated around that table, had the experience of many birthdays and also family debates. The social job of the table communicates that, despite the fact of moving into a smaller apartment, the downsizer has a large table to host friends and guests. Both functions, emotional and social, cannot be fulfilled by a new and smaller table. In essence, making the dining rooms larger helped to increase apartment sales.

The jobs to be done-framework, with a functional, emotional, and social job, offers three helpful perspectives for work space (re-)design. First, it makes clear that spaces and the objects in the space serve more than only a functional job. It is very important for the acceptance and success of space designs to acknowledge and consider these alternative jobs. In particular, most of the users themselves are also not aware of the jobs beyond the functional job and in depth interviews might be needed to find them out (refer to Sect. 3.3 as well). Second, the jobs to be done-framework offers a structure to summarize the analysis of the status quo as well as the different facets of the space redesign. We have even used the three dimensions to adequately brief architects about the jobs that certain areas of the space should fulfill; Fig. 3 refers to an example for a briefing template using the three jobs to be done-types.

Third, the jobs to be done-framework allows breaking down jobs and finding alternatives for each one. For instance, the jobs of a desk can be distinguished as the functional job of holding the keyboard and making a place for writing to be possible, the emotional job of showcasing family pictures, and the social job of being the unofficial meeting point for colleagues. If, during a space redesign, this desk should be changed, we can find adequate alternatives for all three jobs, because we are aware of them. A comfortable chair with a small work surface for the keyboard, and a clipboard to write on, or a small standing desk might fulfill the functional job. Whereas the family pictures could be put on a magnetic board or a trolley and a dedicated area in the office close to the coffee machine fulfill the social outlet.

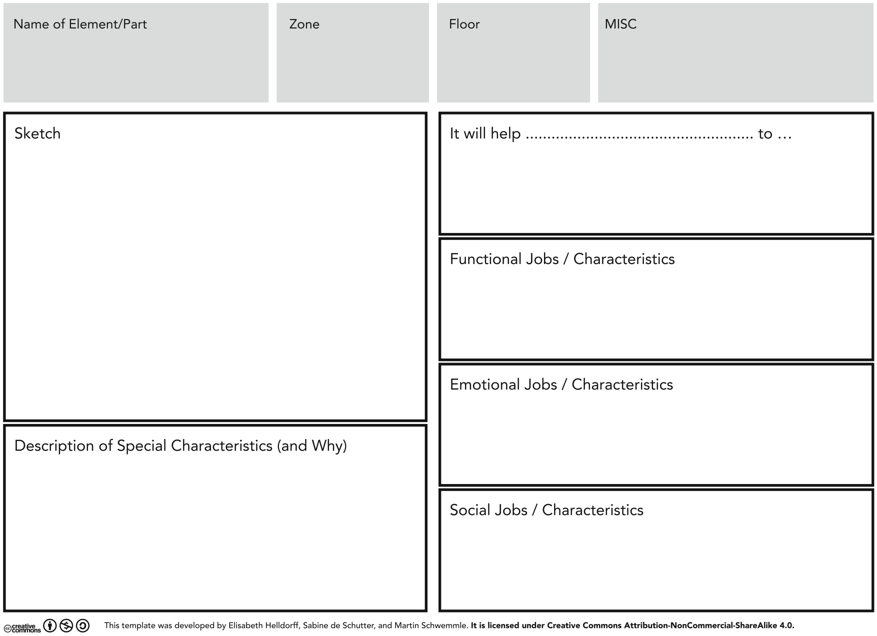


Fig. 3 Documentation template using jobs to be done-framework

In our experience, introducing the jobs to be done-framework and illustrating it through examples supports creating a broader understanding for the needs of employees and also helps them to better reflect their own space usage.

3.2 *User Journeys: Understand Who Does What with Which Tools*

Effective workspace design in organizations highly depends on the people using the space as well as their specific needs in regard to their daily work life. Therefore, it is crucial to understand how different users move around the space, what they do where, and which tools and resources they use. Such an understanding helps to identify not only specific needs of different people within an organization, but also identifies potential touchpoints in terms of frequently used places or, in contrast, the lack of them. Frequently used places are relevant for different reasons: (1) because they most likely work well in terms of fulfilling current needs and (2) they bear a large potential as interaction and connection points among colleagues and become (informal) meeting points. These meeting points enhance overall collaboration as they offer opportunities for coincidental interactions that are fruitful opportunities to exchange knowledge (e.g., Phelps et al. 2012). Time also provides an important dimension of spatial use. Not only is it important to know what places are used for what, but also what time of the day something happens or in which phase a project finds itself in.

Creating a user journey along either a typical working day or a whole project is an effective tool to gain an overview of the needs of employees, their daily tasks and differences or their shared commonalities (e.g., Kumar 2013). For each employee function and/or position, separate user journeys should be created to be compared with each other. Whether to look at only one day or an entire project as a time frame depends on the nature of work. In cases where daily tasks are usually the same, a day already tells a lot. For employees whose daily tasks change often, but overall projects are comparable, this might be a better choice.

In order to create a user journey, we suggest the following steps: Create a list of employees or group employees with similar jobs. Chronologically go through (for example) a full workday to define (1) where employees do each of these activities (places), (2) what they do there (activities) and (3) what they need and use for these activities (resources and tools). Mark the different users, for example by using different coloured post-its. If you order all the information in three rows (places, activities, resources and tools), potential touchpoints between users, that might be positive or harbor a risk for distractions (for example) can easily be defined (please refer to Fig. 4). Distractions might result from different activities, which cause and need different noise levels and take place at the same place or in close proximity to each other. For instance, different employees work at their desk in a large open-floor office, some make phone calls with clients, others work silently and need to focus to prepare an important presentation. In contrast, identifying similar activities taking place at different locations also helps to create new touchpoints between employees by centralizing activities or resources. For instance, everyone uses the coffee machine in the morning, but everybody on different floors. If coffee were only available on one floor, this could naturally become a “morning meeting point.” Furthermore, a comprehensive user journey of different parties within the organization also helps to raise awareness and provides a valuable overview of different functions that the space has to fulfill. These functions could then be summarized and categorized along the jobs to be done-framework.

3.3 Going Beyond: Some Hands-on Exercises

The two preceding sections have shown two important tools to better assess the status quo. However, since most employees are not used to reflecting their usage of space or to making their latent needs explicit, in this section we propose three exercises that can support this process:

– Make latent needs explicit
– Trigger creative solutions by radical spaces
– Find alternatives for furniture.

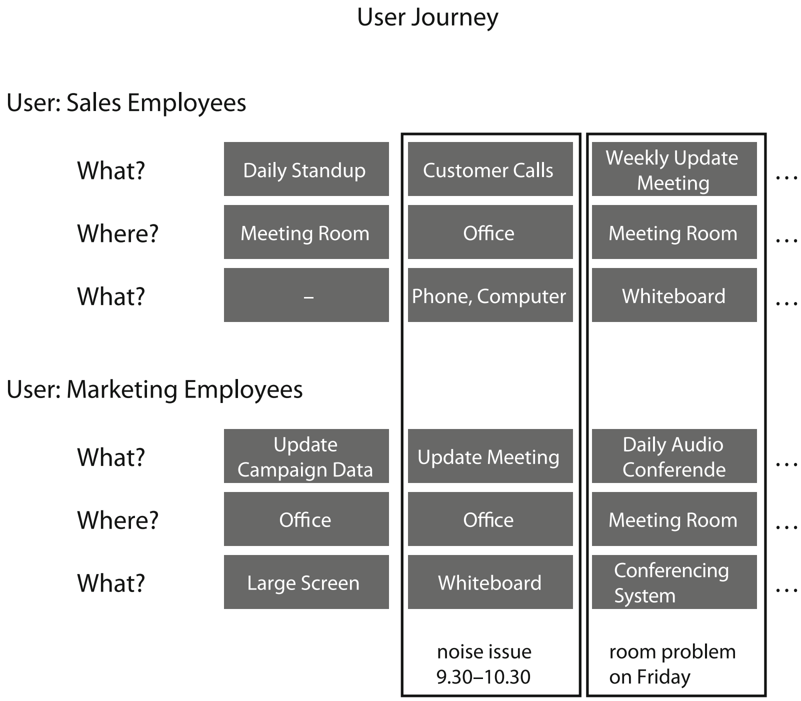


Fig. 4 Example for a user journey and identified conflicts

Make Latent Needs Explicit

Latent needs are "existing but not … manifest" (Oxford Dictionary), which means that people have those needs, but they are not aware of them and therefore cannot articulate them. Making these latent needs explicit is a topic strongly discussed in marketing and management, since it offers companies a way to better satisfy customers' needs and thereby generates a competitive advantage (Narver and Slater 1990). As a consequence, a number of techniques are discussed in literature. We want to share one experience-driven exercise in order to help identify latent needs with regard to work space.

First, start with the specific situation and then use your imagination to think of future possibilities. If you start a conversation by asking people to envision their optimal work place directly, in most cases, you will get unsatisfactory results. In our experience, it is much better to start with the current situation and ask what they like and dislike about it today. Then, one can ask if they have seen other good examples that they would like to have for a work place. And last, you can then ask about wishes regarding the future work place. A technique that also works well during such interviews is offering alternatives as instant solutions or suggestions to problems to get feedback directly and thereby also uncover latent needs. These suggestions do not necessarily have to be realistic, but, following a "what if …"

mode, can also push boundaries to learn about needs and then be translated back into reality.

For instance, during a workshop an employee mentioned that he would like to have some focus time for working on complex situations. His shared his office with many colleagues who often came and went when they had questions, which did not allow for much quiet time. We proposed a quiet space outside the office where he could go. He liked the idea, but hesitated leaving his desk. During the course of the conversation with him, we found out that there were no technology means in place in order that he could easily take his phone with him for urgent calls. Furthermore, when he needed some important documents they would always be on his desk. We next asked about a mobile trolley that could hold these documents and found out that it could be the solution to his problem. The phone issue was then handed over to his manager. To illustrate a more radical suggestion, a further suggestion would be a trolley that could automatically follow the employee and thus allow him instant access to all needed documents and tools wherever he is.

In a similar vein, one employee mentioned the distractions from people coming into her shared office asking her colleagues for advice. Although she was hesitant regarding hot desking, we asked if she could imagine a rotating reception desk so that one of the colleagues in the office could always take incoming questions. With the benefit of less disturbances she would even be open for giving up "her" desk.

To sum it up, as the given examples have illustrated, open conversations with employees that aimed at a deeper understanding of needs and reasons underlying them—without relying on standardized questionnaires or structured interviews—helped to identify employees' latent needs. Only such deeper digging made it possible to identify and solve their issues—instead of only dealing with superficial symptoms. Consequently, this approach allowed the new work place to add value to the work of the employees.

Trigger Creative Solutions With Radical Spaces

A more creative exercise that also helps to open up the potential for creative solutions is the use of radical spaces or environments. For instance, we would present a deck of cards to people that showed well-selected metaphorical pictures of different spaces that have a deep cultural meaning. These included a kitchen, a tent in the mountains, an opera hall, a train station, and a huge construction site. Next, we asked them what they would need in order to work there. The natural settings, the archetypical nature and the deeply-rooted underlying cultural meanings of these spaces often help to make clear what the employees really need to properly do their work. We further asked them to name possible alternatives of current tools or furniture that could also work in their environments. We did this by creating references to these archetypes of spaces. A tent in the mountains, for instance, offers great inspiring views to be creative, but would not make it possible to store a lot of stuff, such as a huge white board or a trolley with material; and it might get dark early resulting in a limited potential work time. Thus, participants in a workshop who picked the "tent in the mountain" as a meaningful archetype for their ideal work space developed a material belt. It held paper, post-its, and pens and would, in combination with a headlight, make it possible to run an ideation session on the mountain peak using the tent itself as a whiteboard. Given the time and nature of the workshop, one can even build simple prototypes based on these ideas to illustrate what is possible.

An alternative way to use the card deck with illustrations of different spaces is to consider them as analogies. For instance, at the beginning of a workshop we handed out a deck of cards with illustrations of many different archetypical spaces and asked participants which of the spaces adequately described their current work space. Doing this exercise in a group already triggers a lot of discussions about the (different) perceptions of the current work space. For instance, participants in a workshop referred to their current workspace as the train station with a lot of commuting strangers. Other participants described their work space with the picture of a kitchen, with the kitchen table as the central node. They would see their work place in the family having lunch at the kitchen table, where the whole team with its old and new (and younger) employees comes together.

Find Alternatives for Furniture

Furniture is usually designed for a specific purpose. In regard to classical office furniture it is commonly understood what tables, chairs etc. are used for or how they support daily work. Yet, it is nevertheless helpful to reflect on the actual purpose that the furniture serves, as this might differ between individual employees. For example, for some a large desk provides space to sketch out processes or to order different tasks in terms of different places on the table. For others, it might mainly fulfill the purpose to store documents and pictures of their loved ones, or it may serve as a (home-)base where a bag and jacket are placed in the morning as one actually moves around the building between meeting rooms. If the purpose and function of a piece of furniture are clear, one can find alternative ways to fulfill this purpose and function. For instance, a desk serving most of the time as a home base can be replaced by a much smaller table without changing the purpose. A reflective exercise to identify needs fulfilled by the particular furniture is to ask yourself:

(1) What do I use the furniture for? (Purpose).
(2) How does it work? (Function).
(3) What do I like (and therefore need) about this specifically? (Feature).

After answering these questions, differentiating the functions along the three jobs—functional, emotional, and social—could be helpful. In the next step think about possible alternatives or ways to re-use existing furniture.

4 The Behavioral and Cultural Components of Space

The third issue we have already introduced at the beginning of this chapter deals with the behavioral component of space. If we consider space as the interaction of spatial elements with a user, a change of space means a change of this interaction and thus, also of the behavior of its users. New behaviors include the need to establish new routines and rituals of behavior. In the following, on one hand, we give illustrations for changing a space mainly through new behaviors and, on the other hand, we talk about the component of cultural change and change management that should (and will) accompany physical space design projects.

4.1 Changing Space Through Changing Behavior

Change of Individual Behaviors

A classical office setup provokes a rather clearly intended use, i.e. a chair to sit on and a table to write on for individual focused work. However, new furniture calls for high investments and also has to fulfill legal standards and regulations, such as fire and general work safety regulations. Besides refurbishing an office and changing the entire interior design, existing furniture can offer new opportunities—if used in a different way. We give four examples of

(1) re-using existing furniture (space hacks),
(2) changing usage patterns,
(3) changing the surroundings, and
(4) changing the social setting.

First, experimenting with and trying out new uses of classical furniture is an easy starting point for changing space. For instance, a mid-height shelf, a coffee table or even a window sill might easily become a perfect standing desk. Or the couch in an office that is usually only used for customer meetings can be simply turned into a place to read reports or even answer e-mails by adding a small table to hold a tablet and some pens. Adding adhesive foil or just writing directly on the doors of a cabinet transforms it into a whiteboard. Such easy re-uses or 'space hacks' offer a variety of possibilities to work differently than usual. They provide a new feeling of work within a known and given setting and thereby also inspire new ways of thinking. Research suggests inspiration to comprise an evocative object ('inspiration by') and a motivating object ('inspiration to') and, thus, space hacks provide possibilities to evoke inspiration by new uses (Thrash and Elliot 2004).

Of course, also more radical space hacks than the mentioned examples are possible: An example would be using the back of a table as board to pin ideas. It could be used by a team during an ideation session, while lying on the floor. This may seem extreme to some readers, however, it was an idea developed and put into practice during one of our space workshops.

Second, space can be changed by changing people's movement patterns in the same physical space. For example, instead of using the coffee machine closest to the office, one could use the coffee machine that is on a different floor. Instead of answering e-mails at your desk, make it a routine to always do e-mails at your high table. Or force yourself to do all your phone calling in a standing position. Employees can themselves also develop or design these different commandments of using a space differently as compared to the usual manner. It is, however important to create a common commitment. Thus, having identified the new routines, employees have to make a commitment to each other that they will indeed follow these new routines (e.g., “I will use a different coffee machine each Monday morning,” or “I will answer my e-mails at the shared high-table desk every Tuesday.”).

Third, existing non-related work environments can also be rediscovered by working at these ‘newly discovered’ places. For example, use the outside area of the office building, which may actually be designed for leisure and breaks, and make it your temporary workplace. If you feel stuck, just grab a paper and pen and go for a walk to reflect on your work. You’ll then be able to write down ideas and thoughts you may have on the go. Or, go to the lobby where you find a lot of traffic; use the canteen in off-peak times to have a cafe-like atmosphere. Next time you walk through your office building, look for potential new work places. You might be astonished how many possibilities are just some steps away. Working from these new places not only provides new inspiration, it also makes possible, more or less, random encounters with other people and thereby generates new thoughts for space re-design.

Fourth, work places are also shaped by the people working around us. You may therefore gain valuable, new perspectives if you surround yourself with new or different people (such as from another department). To this end, ask a colleague from another department who will be gone for a business trip if you can use their desk. Or, initiate a desk swap with other departments so that you can mix up with a lot of different colleagues. You can also join another team’s table for a day or a defined work session during a project, which not only leads to a new surrounding but also makes communication with new people much easier. Again, interacting with people other than the usual colleagues bears the potential of new ideas through discussions with them and might also allow you to see current topics, questions, or challenges through someone else’s eyes.

New Ways of Working

In addition to changing general ritualized behaviors as explained in the previous section, the application of new ways or methods of working in a more structured way and the resulting change in space usage can help to transform spaces. In particular, we will introduce new ways for meetings, changing work modes, and work materials.

One efficient starting point for linking new ways of working with spaces are meetings. Executives spend 23 h per week in meetings, and also regular employees spend a lot of their time in meeting rooms (Porter and Nohria 2018). However, it might be questioned if all these meetings are necessary and efficient and a recent article even asks to for stop the ‘meeting madness’ (Perlow et al. 2017). Why not start with new rules for meetings that are tied to the activities in meeting rooms or “plan a better meeting with Design Thinking” (Bernstein and Ringel 2018)? For instance, all meetings should not primarily be used to inform, but to co-create. These rules can also be supported by some easy measures in the meeting room, such as providing a timer to give each agenda point a planned duration or only having high tables to transform every meeting into a standup meeting, prevent sitting and the resulting meeting fatigue. In addition, only allowing the use of digital devices for the purpose(s) of the meeting to prevent attendees from falling into multitasking and answering e-mails while in attendance. Such behaviors will make everybody question whether all agenda points are really necessary and/or can be fulfilled in different ways. Such measures are relatively simple, but from our own experience we know that by just bringing a timer to a meeting and making the time spent in the meeting visible may result in a completely different tone. In one project, a CEO was so inspired from our first meeting with a timer that he bought his own timer and would bring it to subsequent meetings. Another option, especially to overcome the problem of meeting room scarcity, is to allow meetings to take place in meeting rooms only for confidential content. Otherwise, meetings should be conducted at other places, such as regular office spaces, lounges, or the cafeteria.

A switch between different work modes during the day is important, since body posture can influence how we think and act—a phenomenon that is known as 'embodied cognition' (Shapiro 2014). This switch applies both to different positions—sitting at the desk, standing, sitting on a couch for reading—while working and on breaks. Further recommended is a switch between working in an analogue and in a digital mode. For many knowledge workers, the workday takes place (almost) exclusively in front of a digital screen. However, this is not always necessary and might even prevent a spatial change due to a lack of space for the screen or power. Why not dictate e-mails via a transcription tool from a more relaxed position or draw the concept for a presentation on a sheet of paper and then—once the thought process is done—copy the drawings in a software? In addition, many times in-consuming e-mail 'conversations' could be replaced by a phone call, or even better, a quick face-to-face talk by just walking to the other person's desk.

4.2 Changing a Space as Change Management

The Eight Steps of Change Management

Understanding space as an interaction of users with spatial structures makes clear that a space design project cannot focus exclusively on physical space. It also needs to deal with behavioral changes of the users to establish new, changed, patterns of behavior. In addition, space can be a manifestation of corporate culture and, vice versa, a way to trigger culture and cultural change. For instance, flat hierarchies can be made visible when the manager's office space is used as a shared meeting space while the manager is away at meetings. This should be possible without having to ask for permission every time. Establishing shared offices, where the manager sits with the team, is another such expression. As one workshop participant once summed it up: "every space design project is a change management project." The models used in change management also hold true for space projects. We want to illustrate this notion along John Kotter's eight steps of change management (Kotter 1996) to underscore this conclusion.

1. *Establish a Sense of Urgency*. To trigger change, the buy-in from management is necessary. Recent global studies, for instance Gensler's Workplace Study (2016), illustrate the necessity and potentials of work space redesign, and it is increasingly becoming a topic of employer branding. In addition, framing the topic of work space on a broader level (e.g., in the context of new work) and showing cases of other companies that successfully redesigned space might help raise management's awareness.
2. *Form a Powerful Guiding Coalition*. As this chapter has illustrated, a space redesign project includes a lot of different areas and therefore requires contributions by different people from different departments. Forming a multidisciplinary team that supports the whole space project is extremely helpful. Part of this team could be representatives from management, the employees who will move into the new space, facility management, HR, IT, or unions.
3. *Create a Shared Vision*. Having a clear vision of what the spatial change should achieve and formulating the strategic intent (Moultrie et al. 2007) underlying the space (re-)design project not only helps to communicate the vision to other employees. In this way, the team (see 2) is also supported by having a shared and motivating common goal.
4. *Communicate the Shared Vision*. In communicating the vision, it is clear that the space design project goes beyond an interior architectural project. Basically, sharing some of the thoughts that are mentioned in this chapter will also help employees, or even external audiences, know what the project is all about. In addition to communication via established channels, such as intranet or social media, workshops where employees are actively involved, for instance analyzing the current space usage, also contribute to communication.
5. *Remove Obstacles*. Since every organization and every project is different, the obstacles that can occur also vary greatly. We will therefore not go into further detail here. Nevertheless, it is important to highlight that building projects are often influenced by external factors, such as environmental regulations or even the weather.
6. *Create Short-Term Wins*. Creating quick wins, in particular during longer projects, shows that something is happening and also sustains the motivation of the responsible team. For space design projects, such quick wins could be pop-up spaces that already showcase how the new spaces might look or activities on the construction site. In addition, running pilot projects for one meeting room, and iterating based on these experiences before a full rollout, is often helpful.
7. *Build on the Change*. As we have already noted, space design projects do not stop with the opening of new spaces. Thus, do not stop the change project too early, but collect insights from the first weeks of using the new space and learn—both how to adapt the space and the change management process.
8. *Anchor the Changes in the Revised/New Corporate Culture*. This last step refers to the link between the work space, related work behaviors, and corporate culture. Unfortunately, we have seen many examples of newly designed innovation spaces or creative labs that have been unused some months after their opening. Thus, the

anchoring of the behaviors linked to the new spaces within the corporate culture is really crucial.

5 Discussion

Building on our experiences of space (re-)design projects, we have linked the most prominent pitfalls practitioners regularly encounter with perspectives from theory. Our aim is to achieve a better understanding of what is underlying these pitfalls and which tools and methods can help to overcome them. In particular, we have discussed a broader understanding of work spaces on the basis of an interactionist understanding, the relevance of the assessment of the current space usage, and the necessity of also viewing space design projects from a change management perspective.

In particular, we have introduced an understanding of space as the interaction between spatial structures and space users and thereby broaden a solely physical understanding of space. As our interactionist model of space shows, it is of great importance to consider the organizational context when dealing with (a) work space. Numerous factors create specific requirements for work space (re-)design projects and, like employee needs, can be latent and might have to be explicated. With users becoming a part of the understanding of space, their emotions and perceptions in this regard also come into play. We have introduced psychological ownership and affordances as theoretical foundations to better understand user' emotions and perceptions. The example of the Frankfurt kitchen further gives a new perspective on the question of the radical redesign of space size.

We additionally highlight the relevance of a proper analysis of the status quo and introduce the jobs to be done-framework, user journeys, and the concept of latent needs as useful theoretical frameworks. We elaborate on the concept of latent needs and describe some hands-on tools in order to make them explicit. In addition to considering such changes of perspectives at the beginning of a project, it must be clear that the unpredictable environment will require employees to remain in a situation, but also to adapt to changing conditions.

Last, we discuss the behavioral component of space. On the one hand, we present examples for changing personal behavioral patterns and introduce new ways of working, on the other hand we consider space design projects as organizational change management projects to underline the impact any space (re-)design project can have in establishing an new way of working within your organization.

An open topic to further elaborate is the digitalization of work and the work space itself and how it raises the question of a useful combination of offline and online spaces and tools. On the one hand, digitalization makes physical offices obsolete, since people will be able to collaborate from anywhere. On the other hand, as a counterbalance, physical presence and interaction might become even more relevant.

In general, from a research perspective, this chapter has hopefully identified a number of topics that provide new starting points for further research. For instance, the perspective of space projects as change management begs a deeper analysis and allows a longitudinal case study approach. In addition, the effectiveness and the boundary conditions of the tools we present could be considered during further research.

Acknowledgements The authors want to thank Theresa Weinstein for her valuable comments to earlier versions of the manuscript and the numerous participants of our space workshops for their valuable feedback to our spatial exercises. Additional thanks to Dr. Sharon Nemeth for her copyediting support.

References

Bacevice, P., Burow, L., & Triebner, M. (2016). 7 factors of great office design. *Harvard Business Review*, 2–8.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bernstein, M., & Ringel, R. (2018, February). Plan a better meeting with design thinking. *Harvard Business Review*.

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15,* 181–195. https://doi.org/10.1207/S15326969ECO1502_5.

Christensen, C. M., Hall, T., Dillon, K., & Duncan, D. S. (2016). *Competing against luck*. New York: HarperCollins.

Dawkins, S., Tian, A. W., Newman, A., & Martin, A. (2017). Psychological ownership: A review and research agenda. *Journal of Organizational Behavior, 38,* 163–183. https://doi.org/10.1002/job.2057.

de Jong, C. W. (2017). *Ten principles for good design: Dieter Rams*. Munich, London, New York: Prestel.

Elsbach, K. D., & Pratt, M. G. (2007). The physical environment in organizations. *Academy of Management Annals, 1,* 181–224. https://doi.org/10.1080/078559809.

Fayard, A.-L., & Weeks, J. (2007). Photocopiers and water-coolers: The affordances of informal interaction. *Organization Studies, 28,* 605–634. https://doi.org/10.1177/0170840606068310.

Frederick, C. (1915). *Household engineering*. Chicago: American School of Home Economics.

Gensler. (2016). US Workplace Survey 2016. https://www.gensler.com/research-insight/workplace-surveys/us/2016 (accessed August 2, 2019).

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Hofer, C. (1996). The metaphysics of space-time substantivalism. *Journal of Philosophy, 93,* 5–27. https://doi.org/10.2307/2941016.

Jerram, L. (2011). Cultural geographies kitchen sink dramas: Women, modernity and space in Weimar Germany. *Cultural Geographies, 13,* 538–556. https://doi.org/10.1191/1474474006cgj374oa.

Klooker, M., Nicolai, C., Matzdorf, S., et al. (2016). On creating workspaces for a team of teams: Learnings from a case study. In H. Plattner, C. Meinel, & L. Leifer (Eds.), *Design thinking research* (pp. 67–84). Berlin: Springer.

Kotter, J. (1996). *Leading change*. Brighton: Harvard Business School Press.

Kristensen, T. (2004). The physical context of creativity. *Creativity and Innovation Management, 13,* 89–97.

Kumar, V. (2013). *101 design methods*. Hoboken: Wiley.

Laing, A., Craig, D., & White, A. (2011). High-performance office space. *Harvard Business Review, 89*, 32–33.

Lefebvre, H. (1991). *The Production of Space*. Oxford and Cambridge: Blackwell.

Löw, M. (2008). The constitution of space: The structuration of spaces through the simultaneity of effect and perception. *European Journal Social Theory, 11,* 25–49. https://doi.org/10.1177/1368431007085286.

Moultrie, J., Nilsson, M., Dissel, M., et al. (2007). Innovation spaces: Towards a framework for understanding the role of the physical environment in innovation. *Creativity and Innovation Management*, *16*, 53–65. https://doi.org/10.1111/j.1467-8691.2007.00419.x.

Museum of Modern Art. (2010). *The frankfurt kitchen*. https://www.moma.org/interactives/exhibitions/2010/counter_space/the_frankfurt_kitchen/(accessed January 4, 2019).

Narver, J. C., & Slater, S. F. (1990). The of effect market orientation on business profitability. *Journal of Marketing, 54,* 20–35. https://doi.org/10.2307/1251757.

Noever, P. (Ed.). (1992). *Die Frankfurter Küche: Edition Axel Menges*. Berlin: Ernst, Wilhelm & Sohn.

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Orlikowski, W. J. (2015). Practice in research: Phenomenon, perspective and philosophy. In D. Golsorkhi, L. Rouleau, D. Seidl, & E. Vaara (Eds.), *Cambridge handbook of strategy as practice* (2nd ed., pp. 33–43). Cambridge: Cambridge University Press.

Perlow, L. A., Hadley, C. N., & Eun, E. (2017). Stop the meeting madness. *Harvard Business Review, 20,* 61–69.

Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of Management, 38,* 1115–1166. https://doi.org/10.1177/0149206311432640.

Pierce, J. L., Kostova, T., & Dirks, K. T. (2011). Toward a theory of psychological ownership in organizations. *Academy of Management Review, 26,* 298–310.

Porter, M. E., & Nohria, N. (2018, July–August). How CEOs manage time. *Harvard Business Review*.

Sartre, J.-P. (1956). *Being and nothingness: A phenomenological essay on ontology*. New York: Washington Square Press.

Schön, D. A. (1983). *The reflective practitioner*. New York: Basic Books.

Schwemmle M., Nicolai C., Klooker M., & Weinberg U. (2018). From Place to Space: How to Conceptualize Places for Design Thinking. In: Plattner H., Meinel C., & Leifer L. (Eds) *Design Thinking Research. Understanding Innovation* (pp. 275–298). Springer, Cham. https://doi.org/10.1007/978-3-319-60967-6_14.

Shapiro, L. (Ed.). (2014). *The Routledge handbook of embodied cognition*. Milton Park: Routledge.

Steelcase. (2016). Steelcase global report: Engagement and the global workplace, 1–249.

Thrash, T. M., & Elliot, A. J. (2004). Inspiration: Core characteristics, component processes, antecedents, and function. *Journal of Personality and Social Psychology, 87,* 957–973. https://doi.org/10.1037/0022-3514.87.6.957.

Whittington, R. (2006). Completing the practice turn in strategy research. *Organization Studies, 27,* 613–634. https://doi.org/10.1177/0170840606064101.