ROBERT TSCHARN

INNOVATIVE AND AGE-INCLUSIVE INTERACTION DESIGN WITH IMAGE-SCHEMATIC METAPHORS

Inaugural-Dissertation zur Erlangung der Doktorwürde an der Fakultät für Humanwissenschaften der Julius-Maximilian Universität Würzburg vorgelegt von Robert Tscharn aus Nürnberg

WÜRZBURG 2019

Erstgutachter: Prof. Dr.-Ing. Jörn Hurtienne
Zweitgutachter: Prof. Dr. Harald Reiterer
Tag des Kolloquiums: 14.12.2018

Acknowledgements

This dissertation was conducted at the Chair for Psychological Ergonomics at Julis-Maximilians-Universität Würzburg. I want to take the opportunity to thank all the people who supported me in the last years on this journey. This thesis would not be possible without the guidance and impulses of Jörn Hurtienne, my advisor and – let's be honest – *the* promoter of image-schematic metaphors in HCI. He inspired me not only for the topic of the thesis but also gave me the freedom to explore other topics of HCI. I am grateful for the invaluable feedback on many manuscripts for conferences and this thesis. My thanks also go to Harald Reiter, my second advisor, for sharing his expertise and thoughts on the topic of my thesis. The studies presented in this thesis owe much to the help of my students and colleagues. Judith Scheiderer, Stephan Huber, Armin Winkler, Kristian Baumann, Tina Köhler, Tim Kawohl, Viola Mocke, Sarote Jilski and Dominik Lipp supported me in collecting data for the user studies. I thank especially Philipp Schaper for refreshing (or introducing?) the more delicate details of statistics (and sometimes joy with a "Bilderbuch"-story, a Vietnam-trip or simply a muffin). Also, my other colleagues Nam Ly-Tung, Tobias Grundgeiger and Carolin Wienrich deserve my deep gratitude for their constructive feedback on good or bad ideas. For final comments on the manuscript of the thesis, I thank Mareike Hoffmann, Gudrun Reindl, Johannes Grosser, "Brieli" Brielmaier and Nora Gosch. Finally, I want to thank Diana Löffler. This thesis also would not have been possible without you on my side. From the mid-thesis-crisis until the final stages of writing up, she was a great support and conversation mate about probably design-irrelevant IS-M like WER ZULETZT LACHT IST EIN CIRCLE. I am very happy to have Diana at my side. Finally, I take this opportunity to express profound gratitude from my deep heart to my parents for their continuous support throughout my life.

Abstract

The field of human-computer interaction (HCI) strives for innovative user interfaces. Innovative and novel user interfaces are a challenge for a growing population of older users and endanger older adults to be excluded from an increasingly digital world. This is because older adults often have lower cognitive abilities and little prior experiences with technology. This thesis aims at resolving the tension between innovation and age-inclusiveness by developing user interfaces that can be used regardless of cognitive abilities and technology-dependent prior knowledge.

The method of image-schematic metaphors holds promises for innovative and age-inclusive interaction design. Image-schematic metaphors represent a form of technology-*independent* prior knowledge. They reveal basic mental models and can be gathered in language (e.g. BANK ACCOUNT IS CONTAINER from "I put money INTO my bank account"). Based on a discussion of previous applications of image-schematic metaphors in HCI, the present work derives three empirical research questions regarding image-schematic metaphors for innovative and age-inclusive interaction design.

The first research question addresses the yet untested assumption that younger and older adults overlap in their technology-independent prior knowledge and, therefore, their usage of image-schematic metaphors. In study 1, a total of 41 participants described abstract concepts from the domains of online banking and everyday life. In study 2, ten contextual interviews were conducted. In both studies, younger and older adults showed a substantial overlap of 70% to 75%, indicating that also their mental models overlap substantially.

The second research question addresses the applicability and potential of image-schematic metaphors for innovative design from the perspective of designers. In study 3, 18 student design teams completed an ideation process with either an affinity diagram as the industry standard, image-schematic metaphors or both methods in combination and created paper prototypes. The image-schematic metaphor method alone, but not the combination of both methods, was readily adopted and applied just as a well as the more familiar standard method. In study 4, professional interaction designers created prototypes either with or without image-schematic metaphors. In both studies, the method of image-schematic metaphors was perceived as applicable and creativity stimulating.

The third research question addresses whether designs that explicitly follow image-schematic metaphors are more innovative and age-inclusive regarding differences in cognitive abilities and prior technological knowledge. In two experimental studies (study 5 and 6) involving a total of 54 younger and 53 older adults, prototypes that were designed with image-schematic metaphors were perceived as more innovative compared to those who were designed without image-schematic metaphors. Moreover, the impact of prior technological knowledge on interaction was reduced for prototypes that had been designed with image-schematic metaphors. However, participants' cognitive abilities and age still influenced the interaction significantly.

The present work provides empirical as well as methodological findings that can help to promote the method of image-schematic metaphors in interaction design. As a result of these studies it can be concluded that the image-schematic metaphors are an applicable and effective method for innovative user interfaces that can be used regardless of prior technological knowledge.

Zusammenfassung

Innovative Benutzungsoberflächen sind eines der Hauptziele der Mensch-Computer Interaktion. Diese neuartigen Benutzungsoberflächen sind eine Herausforderung gerade für ältere Benutzer und drohen diese aus der immer digitaleren Welt auszuschließen. Hierbei spielen abnehmende kognitive Fähigkeiten und eine geringere Vorerfahrung mit Technologie eine wichtige Rolle. Diese Arbeit zielt darauf ab, die Spannung zwischen Innovation und Alters-Inklusivität zu verringern und Benutzungsoberflächen zu entwickeln, die unabhängig von kognitiven Fähigkeiten und technologieabhängigem Vorwissen benutzt werden können.

Die Methode der image-schematischen Metaphern verspricht innovative und zugleich alters-inklusives Interaktionsdesign. Image-schematische Metaphern stellen eine technologie*unabhängige* Form von Vorwissen dar. Sie offenbaren grundlegende mentale Modelle und können aus metaphorischer Sprache extrahiert werden (z.B. Bankkonto ist Container ausgehend von "Geld *ein*zahlen"). Die vorliegende Arbeit leitet aus vorangegangen Anwendung von image-schematischen Metaphern im Bereich der Mensch-Computer Interaktion drei empirische Forschungsfragen mit dem Fokus auf innovatives und alters-inklusives Interaktionsdesign ab.

Die erste Forschungsfrage behandelt die bisher ungetestete Annahme, dass junge und ältere Menschen in ihrem technologieunabhängigem Vorwissen und damit auch im Gebrauch image-schematischer Metaphern übereinstimmen. In Studie 1 beschrieben 41 Probanden abstrakte Konzepte in den Bereichen Online Banking und Alltag. In Studie 2 wurden zehn kontextuelle Interviews durchgeführt. In beiden Studien wurde eine Übereinstimmung zwischen 70% und 75% gefunden, was auf eine substantielle Übereinstimmung der mentalen Modelle hinweist.

Die zweite Forschungsfrage zielte auf die Anwendbarkeit und das Potential image-schematischer Metaphern für innovatives Design aus der Perspektive von Designern ab. In Studie 3 durchliefen 18 studentische Designteams einen Ideenfindungsprozess mit Prototypenerstellung, der entweder auf einem Affinity Diagramm als Industriestandard, image-schematischen Metaphern oder beiden Ansätzen in Kombination basierte. Die Methode der image-schematischen Metaphern, aber nicht die Kombination beider Methoden, war ebenso leicht anwendbar wie die bekanntere Standardmethode. In Studie 4 erstellten professionelle Interaktionsdesigner Prototypen mit oder ohne image-schematische Metaphern. In beiden Studien wurde die neue Methode als leicht anwendbar und die Kreativität stimulierend wahrgenommen.

Die dritte Forschungsfrage ging der Frage nach, ob Prototypen, die explizit auf image-schematischen Metaphern basieren, tatsächlich innovativer wahrgenommen werden und alters-inklusiver bezüglich kognitiver Fähigkeiten und Technologievorwissen sind. In zwei experimentellen Studien (Studie 5 und 6) mit insgesamt 54 jüngeren und 53 älteren Menschen wurden Prototypen, die mit image-schematischen Metaphern entwickelt worden waren, als innovativer wahrgenommen als solche, die

nicht explizit mit der neuen Methode entwickelt worden waren. Zudem war der Einfluss von Technologievorwissen auf die Interaktion geringer für Prototypen, die mit image-schematischen Metaphern erstellt worden waren. Der Einfluss von kognitiven Fähigkeiten und Alter auf die Interaktion blieb jedoch signifikant.

Die vorliegende Arbeit liefert sowohl in empirischer als auch methodischer Hinsicht einen Beitrag zur Weiterentwicklug der Methode der image-schematischen Metaphern im Interaktionsdesign. Als Ergebnis dieser Arbeit lässt sich festhalten, dass image-schematische Metaphern eine leicht anwendbare und effektive Methode darstellen, um innovative Benutzungsoberflächen zu entwickeln, die unabhängig von Technologievorwissen benutzt werden können.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

It is easy to dive into theory, or all the way into just practice - but the real interesting work happens between theory and practice. – Erik Meijer (2017) as cited in Matsudaira and Meijer (2017)

We live in an increasingly digital world. Technological advance has gained momentum and innovation pressure seems to be inherent to many of today's societal and economic fields: tradition is no business model. Whether this is a good or bad thing has been subject to discussions for decades (Kurzweil, Richter & Schneider, 1990; Weizenbaum, 1972). Still, the topic of progressing digitalisation in realms that for a long-time were reserved for human execution attracts prominent supporters and opponents also in recent debates (Solon, 2017).

Innovation has become a buzzword positively connoted and "innovative" is a desirable attribute in many domains ranging from entire organisational processes and research programs to single consumer devices. Also in the field of human-computer interaction (HCI), a striving for "innovation" has gained currency (Norman & Verganti, 2014). Innovative user interfaces and creative interactions are of great interest to researchers and practitioners (Kelley, Woren & Kelley, 2013). The HCI community is indeed in an excellent position to provide innovative approaches for the communication between humans and machines to match this need, even though innovations are often driven by technological advances (Hasan & Yu, 2017). However, technological or functional novelty is not the only approach to innovation in HCI. Products that are perceived as innovative are often simply artefacts that satisfy the human's need for stimulation by novelty and challenge (Hassenzahl, 2004). Moreover, this stimulation can stem from other sources than a functional novelty. Importantly, the user interface alone can already lead to perceived innovation through its presentation and visual representations (Radford & Bloch, 2011). Communicating innovation via the user interface is equally important than "functional" innovation: a car that is technically breaking grounds but looks over-the-hill and from the late eighties will not be taken up as an innovative one by the majority of people. The present work focuses on this innovation on the user interface level and not on functional innovation.

Perceived innovation implies perceived newness (Johannessen, Olsen & Lumpkin, 2001; E. M. Rogers, 1995). Thus, the user of an innovative and novel user interface will experience a certain degree of unfamiliarity and will often have to learn the new form of interaction. At the same time, newness and unfamiliarity should not hinder the user to operate it effectively and efficiently while achieving a high satisfaction. In the best case, innovative forms of HCI still build on the users' previous experiences and knowledge. If not, innovative forms of HCI must be learned, which can be a cognitively demanding task.

Learning innovative forms of HCI becomes more difficult over the lifespan (Barnard, Bradley, Hodgson & Lloyd, 2013; Fisk, Rogers, Charness, Czaja & Sharit, 2009). Younger adults already bring a history of experiences with diverse forms of modern HCI and sufficient cognitive resources to cope with innovative HCI. Older adults – usually referring to people older than 60 or 65 years (Vines, Pritchard, Wright & Olivier, 2015) – often can only draw on usually (non-)technological experiences obtained in a less digital environment. Additionally, cognitive abilities statistically decline early in the ageing process (Baltes & Lindenberger, 1997), which further reduces the probability of learning innovative HCI quickly or even successfully (Mitzner et al., 2010). Even though researchers warn about seeing the group of older adults as too homogenous and negative (Vines et al., 2015), statistically they usually face problems that are highly relevant for innovative HCI. Besides perceptual and physical-motor impairments, they must cope with lacking prior technological knowledge (Fisk et al., 2009; Hurtienne, Horn, Langdon & Clarkson, 2013) and cognitive decline (Salthouse, 2010) rendering exclusion from innovative HCI a severe threat. Symptoms of these challenges while interacting with unfamiliar new technology can be found in higher levels of computer anxiety (Marquié, Jourdan-Boddaert & Huet, 2002), a behaviour of error-avoidance instead of trial-and-error (Barnard et al., 2013) and statistically poorer performance when interacting with technology (Reddy, Blackler, Popovic & Mahar, 2014).

The relevance of the group of older adults is increasing due to demographic change: in 2017, 13 percent of the world population was older than 60 years, and this percentage is growing on average with 3 percent per year (United Nations & Affairs, 2017). This user group must be taken into account when designing innovative HCI. Contrary to public stereotypes (Wandke, Sengpiel & Sönksen, 2012), older adults are not aversive against technology per se and often see useful applications for innovative technologies (Mitzner et al., 2010). Despite this fact, a digital and technological gap between age groups and exclusion of older adults from modern technology has been observed (Czaja et al., 2006; C. Lee & Coughlin, 2015), placing an obstacle for older adults for participating in today's digitalised societies. Unfortunately, the HCI community seems to perceive the digital divide as a cohort effect and adopts the myth of "Just Wait and See: Future Generations of Older People Will Use Computers without Problems" (Wandke et al., 2012, p. 566).

Differences between younger and older adults occur naturally. But resulting digital exclusion is human-made and depends on the age-inclusiveness of interaction design. Perceived ease of use is regarded as a critical factor for technology access and adoption (Davis, 1989; C. Lee & Coughlin, 2015). A variety of guidelines and design recommendations (such as using comprehensible layouts) have been proposed to achieve age-inclusiveness in HCI, which is to enable users to interact with interfaces regardless of their age (Barnard et al., 2013; Fisk et al., 2009; Nurgalieva, Laconich, Baez, Casati & Marchese, 2017). However, current approaches to age-inclusive HCI like guidelines face several limitations.

First, they restrict the degree of innovation by narrowing down possible solutions to ones that are in line with guidelines. Little room remains for creativity, and interaction designers often do not adopt guidelines in the field (Nurgalieva et al., 2017). Well-intended recommendations often cannot find their way to interaction designers, reducing the real-world impact of guidelines. Second, guidelines rarely focus on prior technological knowledge and cognitive decline, two factors that have been shown to have a high impact on interacting with unfamiliar interfaces (Blackler & Hurtienne, 2007; Reddy et al., 2014). Guidelines and frameworks that focus on these two crucial challenges are often constrained to general recommendations that provide little guidance in practice (Nurgalieva et al., 2017). Third and most importantly, existing guidelines and frameworks for interfaces that can be used by older adults usually aim at providing separate user interfaces for younger and older or technology-experienced and inexperienced persons (Reddy et al., 2014; Sengpiel, 2016). This further enlarges the gap between younger and older adults because older adults are blocked from experiencing the same technology as younger adults. This digital exclusion prevents older adults from building comparable technological prior knowledge as younger adults.

Avoiding a technological divide between younger and older adults has received increased attention amongst academics and practitioners under the label *age-inclusive design*. Following the philosophy of universal design (J. Johnson & Finn, 2017; Story, Mueller & Mace, 1998) and inclusive design (Langdon, Lazar, Heylighen & Dong, 2014), users with different age-related capabilities and constraints are considered as one target population of one-size-fits-all HCI. User-Centered Design (e.g. Holtzblatt & Beyer, 2017) is a powerful approach for age-inclusive design. It can improve the development process by including users of different ages while still allowing for enough room for innovative ideas. However, it still lacks concrete guidance on the interface level, a general problem most guidelines also face.

1.1 Underlying question of this thesis

Taken together, the field of HCI steadily strives for innovation which perpetually increases the digital divide between younger and older adults. Age-inclusive interaction design aims at providing HCI that can be used by users regardless of their age. The focus of the present work is on two facets of age-inclusiveness: cognitive abilities and prior technological knowledge. Existing methodologies suffer from severe limitations when innovation and age-inclusiveness are to be combined in one design process. How can we designing for innovation and age-inclusiveness at the same time?

In this thesis, the answer to this question grounds in a framework for interaction design which is inspired by cognitive linguistics (M. Johnson, 1987). So-called *image-schematic metaphors* (e.g. BANK ACCOUNT IS A CONTAINER) represent a form of technology-independent and thus age-independent prior knowledge that can be leveraged in HCI. At the same time, image-schematic metaphors have been proposed to stimulate interaction designers to create innovative designs. The focus of the present work is to use image-schematic metaphors as a tool for innovative and at the same time age-inclusive HCI (Hurtienne, 2017a).

One efficient window to this form of technology-independent prior knowledge is found in natural language. Language analysis has been promoted as a useful tool for unearthing mental representations of abstract concepts (Asikhia, Setchi, Hicks & Walters, 2015; Casasanto & Boroditsky, 2008). Cognitive linguistics reveals how image-schematic metaphors could be grounded in everyday experiences an, therefore, potentially age-independent (M. Johnson, 1987; Lakoff & Johnson, 1980). Extracting image-schematic metaphors from language and using them as guidance for interaction design has been shown to enable designers to successfully match the system model to their users' mental representations of abstract concepts (Hurtienne, 2017a). However, understanding image-schematic metaphors as a window to age-independent mental representations is based on the – yet untested – assumption that younger and older adults show the same image-schematic metaphors in their language.

The potential of image-schematic metaphors for innovative HCI that is still intuitive to use has been shown (Hurtienne, 2011; Winkler et al., 2016). Case-studies show that it is possible to integrate image-schematic metaphors into design processes like User-Centred Design (Hurtienne, Klöckner, Diefenbach, Nass & Maier, 2015; Löffler, Hess, Maier, Hurtienne & Schmitt, 2013). However, the benefits of the new method can only scale when it is comprehensive and accepted by interaction designers. Even though the feedback of previous work was positive (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013), it remains unclear how comprehensive and innovation stimulating image-schematic metaphors are in comparison to other design processes.

By designing with image-schematic metaphors, interaction designers should be inspired to create innovative user interfaces. At the same time, these user interfaces should be in line with age-independent mental representations and, therefore, decrease the impact of age-related differences in prior technological knowledge and cognitive abilities on HCI. Even though previous results addressing this tension are promising (Hurtienne, Klöckner et al., 2015), an experimental evaluation of multiple prototypes that are based or not on image-schematic metaphors is still missing in the literature.

In sum, given the innovation-pressure in the HCI-community and the importance of age-inclusive interaction, image-schematic metaphors represent a promising method for addressing both needs. The goal of the present work is to develop the method of image-schematic metaphors further and guide its application for innovative and age-inclusive interaction design. The following three research questions were identified in the literature, and each is investigated by two empirical studies in the present work:

1. The theoretical assumption must be validated that mental representations of abstract concepts – elicited by comparing the use of image-schematic metaphors – overlap between younger and older adults and thus really can serve as an age-independent form of prior knowledge.

2. The integration of image-schematic metaphors into the interaction design process has been reported in several projects. However, evidence for the comprehensibility of the new method and the potential for innovative interaction design in comparison to other design processes is still missing.

3. Evidence for the promises that HCI that are based on image-schematic metaphors are more innovative and age-inclusive than for HCI that is not explicitly based on image-schematic metaphors is still sparse. Empirical experiments need to support and provide empirical data for these claims.

The contribution of this thesis is twofold: empirical as well as of methodological. On the one hand, it provides empirical evidence to fill the stated research gaps that – until now – disrupt argumentation lines for innovative and age-inclusive design based on image-schematic metaphors. On the other hand, it derives concrete methodological recommendations from the studies to facilitate the application of image-schematic metaphors in future projects.

1.2 Overview of the thesis

The thesis is structured as follows. In chapter 2, the literature on age-related differences and their impact on HCI are summarised. The emphasis is on the role of age-differences in cognitive abilities and prior technological knowledge. The need for age-inclusive interaction design and current design methods are described.

Chapter 3 reviews the concept of innovation. Also, contemporary design processes, techniques and recommendations for innovative design are described. Shortcomings of existing approaches are summarised and the challenge for innovative and at the same time age-inclusive interaction is discussed.

Chapter 4 summarises the method of image-schematic metaphors for interaction design. The theoretical foundations for image schemas and image-schematic metaphors are laid, and their promises for innovative and age-inclusive interaction design are derived. Prior work on image-schematic metaphors in interaction design is reviewed, and similarities, as well as differences in existing applications of this method, are extracted. The method of image-schematic metaphors is discussed regarding its potential for innovative and age-inclusive interaction design, resulting in a set of open research questions that have to be answered before recommending the usage of this method. Finally, the research questions of the present work are made explicit.

Chapter 5 focuses on the overlap of different age groups in their mental representations. It describes two empirical studies (studies 1 and 2) that both focus on image-schematic metaphors in the spoken language of younger and older adults. Similarities and differences between the two age groups are highlighted, and the implications for image-schematic metaphors for age-inclusive interaction design are discussed. Insights for the extraction process of image-schematic metaphors from spoken language are derived, and recommendations for future projects are formulated.

Chapter 6 focuses on the designers' perspective and describes two empirical studies that integrate image-schematic metaphors in the interaction design process (studies 3 and 4). Insights from the design process with image-schematic metaphors in both a standardised experiment as well as a study with professional interaction designers in a company are summarised. From this, recommendations on how to efficiently incorporate image-schematic metaphors into existing work procedures – specifically into User-Centred Design processes – are extracted and compared to recommendations of previous research in this field.

Chapter 7 sheds light on the question whether image-schematic metaphors facilitate innovative and age-inclusive interaction design. It describes two empirical studies (studies 5 and 6) that try to validate these promised benefits. Prototypes that were explicitly designed with image-schematic metaphors are compared to baseline-prototypes that were either designed during a standard design process (study 5) or based on industry standards (study 6). The results shed light on the impact of image-schematic metaphors in interaction design on perceived innovation and age-inclusiveness regarding differences in prior technological knowledge and cognitive abilities.

Chapter 8 finally concludes on the main findings of the present work and their limitations. It recapitulates the insights from the six studies in the light of the literature. Additionally, a general procedure for interaction design with image-schematic metaphors is described that will be useful in future applications of the method. Possible next steps and an outlook for future work close the thesis.

Chapter 2

Designing for Age-Inclusive User Interfaces

In the public perception, older adults are often seen as a technology-aversive user group that is reluctant in keeping pace with technological advancement (Herstatt, Kohlbacher & Bauer, 2011). Stereotypes of older adults focus mostly on the negative aspects of ageing but – both in academia and the customer market – technology is necessary for counteracting these negative aspects (Vines et al., 2015). For example, communication technologies can enhance the connectedness of older adults with their peer-groups and families to cut the threat of loneliness, but older adults' adoption of these technologies drags behind that of younger adults (Hope, Schwaba & Piper, 2014). However, stereotypes (like the "lonely old adult") often do not hold true, and sometimes modern communication devices are merely cast aside to keep one's private space without interruptions by calling family members (Bailey & Sheehan, 2009).

However, older adults are an interesting and highly relevant user group for the field of HCI. From an HCI perspective, age-related changes in perception, motor skills, cognition, attitudes towards technology as well as a multitude of other physiological and psychological characteristics provide challenges for interaction design to be solved (Barnard et al., 2013; Coleman, Clarkson, Dong & Cassim, 2016; Fisk et al., 2009). From an economic perspective, older adults often have the financial background to acquire costly technology as long as they perceive them as useful (Mitzner et al., 2010). Moreover, from a social perspective, they are a growing group and an increasing proportion of western societies is to be considered as "old" (Pew Research Center, 2017; Zimmer & McDaniel, 2016). While in the year 2015, 27.6% of the German population was over 60 years old, it is estimated that in 2030 this number will increase up to 39.3% (United Nations & Affairs, 2017). This trend also applies to the world population in general with 12.3% in 2015 and 21.5% in 2030 (United Nations & Affairs, 2017).

Even though chronological age often serves as the primary indicator for an individual falling into the group of older adults, a general and universally accepted definition of “older adult” is still lacking. In all of their reports, the United Nations consistently apply a threshold of 60 years (United Nations & Affairs, 2017) and large-scale research programs adopt this recommendation (such as the follow-up study of the ’Berlin Age Study’, Bertram et al., 2014). Still, some argue that the line between younger and older adults should be drawn at 65 years (Boot et al., 2013; Fisk et al., 2009; Vines et al., 2015). In the field of HCI, thresholds are often more flexible: Dickinson, Arnott and Prior (2007) reviewed literature on older adults starting from ”60-65 years”, Romano Bergstrom, Olmsted-Hawala and Jans (2013), Ziefle and Bay (2005) and Hanson (2011) included older adults over 50 years in their work, and Blackler, Mahar and Popovic (2010) used the term older adults even for adults only older than 40 years. Wagner, Hassanein and Head (2010) also criticised that thresholds among younger and older adults range from 40 to 75. Even though recruitment of samples of older adults satisfying thresholds of 60 or 65 years is difficult (Dee & Hanson, 2014), practicality and validity must be balanced when conducting studies with this target population. Additionally, a one-size-fits-all threshold based solely on chronological age is a poor indicator of age-related changes.

Importantly, chronological age is usually used only as a proxy-variable for underlying age-related changes and age-related changes become more likely with chronological age (Park & Festini, 2016; Salthouse, 2010). However, if the group of older adults is too young, these age-related changes cannot be assumed automatically. Researchers, therefore, should include direct measurements of consequences of the ageing process for each sample (Blackler, Popovic, Mahar, Reddy & Lawry, 2012; Vines et al., 2015) instead of relying on well-documented correlations between age and age-related changes (Fisk et al., 2009; Neves, Franz, Munteanu, Baecker & Ngo, 2015; Ziefle & Bay, 2005). Therefore, the work presented in this thesis will aim at directly measuring consequences of ageing.

Since ageing is a phenomenon entailing a variety of age-dependent changes in different areas, a clear chronological threshold is not ideal. Valid age-thresholds are likely to be dependent on the context and research question. Instead of assuming age-related differences (e.g. sensorimotoric or cognitive) for different chronological age groups, direct measurement of these differences are more appropriate. In line with this, some researchers argue that studies should instead measure underlying age-related changes (e.g., cognitive performance, auditory thresholds), instead of relying only on chronological age itself (Reddy et al., 2014; Vines et al., 2015), because older adults represent a very heterogeneous user group, even more than younger adults (Fisk et al., 2009). On the one hand, it is well established that some abilities decline with age (e.g. performance in a divided attention task) while some remain intact until old age (e.g. procedural memory, Mynatt, Essa & Rogers, 2000; Salthouse, 2010). On the other hand, the extent of decline differs between people and empirical evidence as well as anecdotal experiences show that older adults are not universally affected by age-related decline (Barnard et al., 2013; Czaja & Lee, 2007; Park & Festini, 2016). Finally, exposure to technology is not distributed equally within the group of older adults and the "digital inequality" in this user group has not vanished in the last years (Hargittai & Dobransky, 2017).

Taken together, most older adults face some changes that imply challenges for interaction with modern technology. But chronological age itself is a bad proxy for declining abilities and, therefore, the actual level of abilities should be reported in research. This chapter, therefore, aims at providing a theoretical context of the user population by summarising critical age-related changes and stabilities that are relevant for the field of HCI and frame the underlying problems of this thesis. First, general sensorimotoric ageing is briefly described to give a basic understanding of well-established characteristics of older adults, even though not the focus of this work (section 2.1). Second, essential theories on cognitive ageing, mechanisms and their influence on cognitive abilities are reviewed regarding their impact on the interaction with technology (section 2.2) Third, research on how older adults adopt new technology and the lack of prior technological knowledge is reviewed (section 2.3). Fourth, current approaches to interaction design for age-inclusiveness are described (section 2.4). Finally, the two major factors for this thesis affecting HCI – cognitive ageing and a lack of prior technological exposure and knowledge – are summarised to make explicit the motivation of this thesis (section 2.5).

2.1 Sensorimotoric Ageing

Most prominently, ageing affects sensorimotoric functioning including vision and hearing impairments and decreases in motor skills (see further the German projects BASE and DEGS, Baltes, Mayer, Helmchen & Steinhagen-Thiessen, 2010; Kurth, 2013). As described above, no clear thresholds exist for the onset of universal sensorimotor ageing, but research shows that decline usually starts earlier than many people think (Baltes & Lindenberger, 1997). To provide a better understanding of the user group of older adults, sensorimotoric changes are summarised in the following section.

Vision impairments can start as soon as with 40 years (Fozard, Gordon-Salant, Birren & Schaie, 2001) and interaction designers have been urged to adapt their interfaces' visual presentation to the abilities of older adults (Brajnik, Yesilada & Harper, 2011; Hawthorn, 2000; Romano Bergstrom, Olmsted-Hawala & Bergstrom, 2016). Research in the area of vision and ageing is mature and a variety of recommendations for HCI can be derived from them. Especially Presbyopia - a reduced field vision and incapability to focus on near objects - has been early reported to play a significant role in vision-related problems of older adults with screen-based interfaces (Czaja & Sharit, 1998; Romano Bergstrom et al., 2016). Additionally, visual contrast sensitivity, as well as abilities in colour vision, seem to decline with age mainly in green and blue wavelength (Hanson, 2011; Stuart-Hamilton, 2002). Furthermore, the often-used white background of websites implies challenges due to glare, especially for older adults (Lopes, 2001). Compared to younger adults, older adults are also less likely to perceive the periphery of websites, where task-relevant status information and additional interactive elements are often located (Romano Bergstrom et al., 2016).

Besides vision-related problems, auditory impairments hamper the interaction between older adults and technologies. While already 20% of all adults between 45 and 54 years report hearing impairments, this percentage increases further to 75% of all adults between 75 and 79 years (Hawthorn, 2000). Especially Presbycusis, the bilateral decline of the ability to hear high frequencies is an impairment that one-third of adults above 65 years are estimated to suffer from (Pichora-Fuller, Schneider, Benson, Hamstra & Storzer, 2006). In other words, higher frequencies are less likely to be perceived by older adults compared to younger adults. These are frequencies that are most important for speech and conversation. As in vision-related challenges, older adults seem to be able to compensate their declining hearing abilities. Even when unclear language is not understood in standardised trials, using context information in daily situations and knowledge about language patterns (e.g., syntax) allow for predicting a most likely best guess that can be assumed in conversation (D. Burke & Shafto, 2008; Fozard et al., 2001).

Finally, also motor-related problems are common among older adults. With age, slower reaction times must be expected, especially in complex tasks (Hawthorn, 2000; Light & Spirduso, 1990), fine-motor-coordination is impaired (Walker, Philbin & Fisk, 1997) and small interaction elements on a tablet are more difficult to use for older than for younger adults (Barnard et al., 2013; Nicolau, Guerreiro, Jorge & Gonzalves, 2014). However, these impairments cannot be found in all older adults (Hughes, Done & Young, 2011) and the performance in reaction tasks has been shown to be dependent on how unfamiliar, interesting, and complex the task is perceived by a participant (Vercruyssen, 1997). Finally, regarding speed-accuracy trade-offs, older adults tend to favour accuracy over speed, which might at least partly explain often reported slower reaction times (Goggin & Stelmach, 1990).

Taken together, older adults - at least statistically - experience several impairments in sensorimotoric abilities. The next section focuses on another crucial type of abilities that severely affects HCI and is highly relevant to the work reported in this thesis: cognitive ageing.

2.2 Cognitive Ageing

Besides sensorimotoric functions, ageing is closely linked to changes in cognitive abilities. More specifically, performance in standardized cognitive tests has been found to decrease with age both in longitudinal and cross-sectional studies (Czaja et al., 2006; Park & Festini, 2016; Salthouse, 2010). It is important to note that cognitive decline in older adults rarely implies severe cognitive impairments such as dementia. Its prevalence is worldwide approximately 6% for people over 60 years. However, it is expected that over 65.7 million people will live in 2030 with a form of dementia (Prince et al., 2013). But contrary to this severe form of cognitive impairment, mild impairments (measured by a decreasing performance in standard cognitive assessments) are the most frequent occurrences of cognitive decline (Craik & Salthouse, 2008).

Many cognitive abilities remain intact during ageing and it is important to emphasise that not all adults and all functions are affected by cognitive decline. Still, the probability of facing a variety of cognitive impairments increases with a person's age. In this section, literature is reviewed focusing on older adults' mild cognitive impairments that are to be regarded as usual during the ageing process (and thus are not classified as a disease like dementia but called cognitive decline).

2.2.1 Theories of cognitive ageing

Cognitive decline subsumes a set of different cognitive functions and abilities. Most prominently, the Catell-Horn-Catell (CHC) Theory of Human Intelligence distinguishes between the fluid and crystalline intelligence (Cattell, 1963). In its essence, fluid intelligence describes basic cognitive functions that allow to solve new problems without relying on knowledge obtained in the past. Crystalline intelligence on the other hand refers to knowledge about the world a human has obtained in the past and which can facilitate to cope with the environment. For example, the process of learning a new language requires fluid intelligence to analyse and understand its mechanisms, while the outcome of this process (having command over the new language, vocabulary) is crystalline intelligence – knowledge that has accumulated over the years through learning. While accumulated crystalline intelligence seems to be less affected by age (Copeland, Bies-Hernandez & Gunawan, 2016), fluid intelligence is at the core of cognitive abilities. Several theories have been proposed to further explain the phenomena of cognitive decline in fluid intelligence (D. Burke & Shafto, 2008; Park & Festini, 2016). This section, therefore, provides an overview of relevant (traditional and more recent) frameworks that serve as the theoretical foundation for valid and reliable measures of cognitive abilities in older adults. In the following, the most important theories will be presented.

As a first group, **Resource Theories** try to explain the worse performance of older adults by "declining resources" in general (Craik & Lockhart, 1972). They argue that the human capacity for processing information is limited and that this pool of resources (e.g. attention, "mental energy") has to be distributed to a variety of cognitive processes. Since older adults are assumed to have fewer resources, they reach the point of no available free resources earlier compared to younger adults (Craik & Byrd, 1982). This explanation of cognitive decline has been criticised, because no measurements of these "resources" have been proposed. Salthouse and Craik (2000) underline the problem of circular reasoning because "the same empirical results that are 'explained' by reduced resources also serve as the primary evidence for inferring the existence of an age-related reduction of resources" (p. 690). Moreover, resource theories cannot explain all effects of cognitive ageing (Park & Festini, 2016). For example, older adults benefit more from top-down context effects and meaningful instead of artificial stimuli than younger adults. Also, older adults seem to rely more on building mental models during text recognition compared to younger adults (Smiler, Gagne & Stine-Morrow, 2003).

The second group of theories on cognitive decline is based on the finding that when tasks include a speed element, performance usually decreases with age. Most prominently, **General Slowing Theories** or also Speed of Processing Theories (Birren, Woods & Williams, 1980; Salthouse, 1996) thus assume that age is accompanied by a decreasing processing speed in general, which in turn affects a variety of other resources such as cognitive operations speed, the ability of inhibition, working memory capacity. Salthouse (1996) distinguishes two different mechanisms that cause cognitive decline: a) older adults *take longer* to process complex information because prior early information processing steps are already slower and b) they can only *exploit less* task-relevant information as younger adults in complex decision-making, again because results from early information steps are not finished yet and cannot be considered for higher-level processing. Over the years, Salthouse could show that measures of speed could explain most of the variance of cognitive tasks (Salthouse, 1985, 1996; Salthouse & Babcock, 1991). Strong support for this theory also comes from the fact that perceptual and motor speed correlate significantly with cognitive speed and speed in general (Salthouse, 1985). However, despite the early assumption that all cognitive functions would decline to a similar extent (Birren et al., 1980; Cerella, 1985), modern theories also take into account findings of function-dependent age-related declines (Craik & Salthouse, 2008; Park & Festini, 2016).

A third group, **Working Memory Theories** focuses on working memory as the primary origin for age-related performance (Baddeley, 1986, 2001). Memory, in general, is usually grouped into four different modules: 1) working memory, 2) episodic memory, 3) semantic memory, and 4) procedural memory (Fisk et al., 2009). Episodic, semantic, and procedural memory are part of the long-term memory and seem to mostly remain intact, even though retrieval becomes slower (Bäckman, Small & Wahlin, 2001; Howard & Howard, 1997). Interestingly, older adults seem to rely even more on semantic knowledge (e.g., focusing on a deep structure of situation models) compared to younger adults (Ackerman & Rolfhus, 1999). This is in line with findings that, during information uptake, older adults focus more on extracting semantic information and building situational and mental models compared to younger adults who are excellent in recalling exact words but do not seem to integrate them into profound situational models as older adults do (Radvansky & Dijkstra, 2007). Older adults seem to use stable mental models to represent complex information. For example, in the form of situation models, even though general cognitive abilities declines with age, older adults do well or even better than younger adults when they have to make inferences from a rich and complex situation (Shake, Noh & Stine-Morrow, 2009). Bäckman et al. (2001) and Salthouse (2010) also emphasised the fact that contrary to fluid intelligence, crystalline intelligence (i.e., accumulated and stabilised knowledge about the world) does not decrease in older adults, at least not until very old age. In sum, semantic memory – that is interconnected knowledge and representations about the world – is assumed to remain more or less intact during the ageing process and older adults form new situational models to structure and integrate their tasks into their prior knowledge more thoroughly than younger adults (Copeland et al., 2016).

Contrary to this, working memory is assumed to be more and more impaired during ageing (Chevalier, Dommes & Martins, 2012; Kowtko, 2012; Salthouse, 2010). Working memory is usually seen responsible for keeping information available for processing or transferring to more durable forms of memory (e.g., long-term). In fact, the transfer from working memory to short-term memory has been shown to be less effective in older adults (Salthouse, 1994), which at least partially can explain smaller memory spans for older adults (Salthouse & Babcock, 1991). However, also long-term information retrieval has been shown to be affected by ageing which could be explained by an inefficient transfer from long-term to working memory (Dijkstra & Misirlisoy, 2009). Further evidence for a cognitive decline on account of working memory comes from cognitive linguistics, where experiments show that older adults tend to adapt their process of syntax generation to a decreasing working memory capacity and simplify their syntax in natural speech (D. Burke & Shafto, 2008). Still, older adults seem to be able to cope with decreasing working memory. Brebion, Smith and Ehrlich (1997) report on strategies older adults exploit to effectively minimise the influence of smaller working memory on performance especially during simple tasks. Since working memory, involving temporary storage and manipulation of information (Baddeley & Logie, 1999), is seen as a core ability for learning and fluid intelligence (Baddeley & Hitch, 1974; Kyllonen & Christal, 1990), an age-dependent decline in working memory can explain a variety of age-related phenomena and is seen today as the explanation of a more general cognitive decline (Chevalier et al., 2012; Czaja et al., 2006; McNab et al., 2015). Developed around the same time as Salthouse's General Slowing Theories, Working Memory Theories of cognitive ageing suggest a bottleneck in older adults' working memory capacity (visuospatial and verbal) and a declining central executive that controls this working memory (Baddeley & Hitch, 1974) to account for most cognitive deficits. Even though only a medium correlation between working memory and processing speed have been reported in literature (Fry & Hale, 2000), working memory seems to be also a good predictor for fluid intelligence. Instead of using processing speed as the principal resource, working memory is seen as an excellent way to measure resource deficits and explain age-related cognitive decline through older adults poorer working memory compared to younger adults (Wingfield, Stine, Lahar & Aberdeen, 1988).

Finally, age-related cognitive decline has been suggested to be linked to a steady decrease in the ability of inhibition of irrelevant information by older adults. **Inhibition Deficit Theories** thus focus on findings that older adults are more prone to being distracted by external and task-irrelevant stimuli (Copeland et al., 2016; Hasher & Zacks, 1988; McNab et al., 2015). For example, older adults experience difficulties in staying focused on a primary task when they have to ignore available but irrelevant information (Rabitt, 1965) and are less effective in inhibiting irrelevant or distracting information and thoughts (Hasher, Stoltzfus, Zacks & Rypma, 1991). They are also more prone to distracting stimuli during visual reading (Connelly, Hasher & Zacks, 1991) and during auditive listening (Hasher, Lustig & Zacks, 2008). Since inhibitory deficits play a significant role in different areas of daily life and interaction with the human's environment, one mechanism that could link this theory to resource theories is that the older adults' inability to filter irrelevant information creates a "clutter" in the working memory and slows down information processing (D. Burke & Shafto, 2008). This approach has been very stimulating for the research community and is still perceived as playing a major role in cognitive ageing research (Park & Festini, 2016).

Also, other theories on cognitive ageing have been put forward such as Degraded Signal Theory (Schneider, Daneman & Murphy, 2005) attributing all cognitive decline to poorer sensory performance or Transmission Deficit Theory (Thornton & Light, 2006) focusing on language and mental representations of knowledge where semantic connections weaken during the ageing process. But these have been criticised or not as broadly applicable as the here presented frameworks and theories.

In sum, cognitive decline is a stable phenomenon that has been subject to basic research for a long time and many theories are available for describing, explaining and predicting cognitive abilities in older adults. Today, a deficit in working memory and an inability of inhibiting irrelevant (internal or external) information can be regarded as the two major cognitive mechanisms that underlie changes in cognitive abilities. However, a combination between different factors (working memory, processing speed, inhibition ability) is likely and no single factor can perfectly explain all aspects of cognitive decline. The fundamental constructs of the here described theories seamlessly translate into valid and reliable tools for measuring cognitive abilities. Importantly, the following measurements cover different aspects of fluid intelligence.

2.2.2 Measurements of cognitive ageing

A variety of instruments is available for assessing the diverse facets of cognitive performance and thus for measuring the individual extent of cognitive decline (Czaja et al., 2006; Wechsler, 1997). Often developed for diagnostic purposes (e.g. in medical settings and disease recognition), researchers can choose from a diversity of cognitive testing tools. This section provides examples of widely applied and theoretically funded tests that allow the quantification of cognitive ageing. The selection of tests is restricted to three criteria: first, since the focus of this work is on older adults, tests that cannot be reliably applied to and compared between participants from different age groups are not discussed here. Second, only tests appropriate for measuring cognitive ageing are summarised and tests focusing solely on perceptual speed, psychomotor speed, and crystalline intelligence are not covered here (e.g., K. H. Schmidt

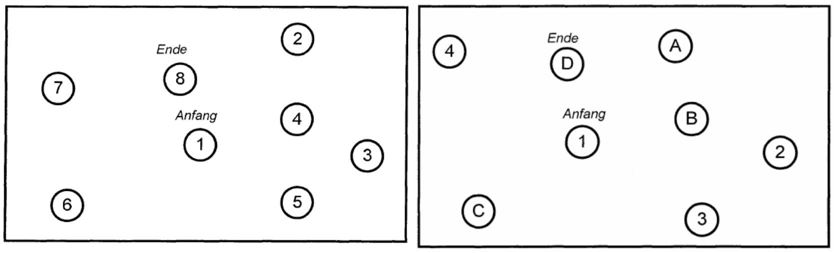


Figure 2.1: Training sheets for the TMT-A (left, connecting only digits) and TMT-B (right, alternately connecting digits and letters).

& Metzler, 1992). Third, only tests that are of practical relevance for HCI-research were included and instruments that are not feasible (ease of administration, data interpretation) are not discussed here.

First, the **Trail-Making-Test (TMT)** with two parts (A and B) is one of the most pervasive cognitive tests in clinical testing (Rabin, Barr & Burton, 2005). Originally developed for detecting brain damages (Reitan, 1958), it has become a reliable and normed instrument for assessing cognitive abilities. In both parts, participants are presented with a paper sheet that contains randomly distributed, but consecutive letters or numbers. In part A, the participant has to draw lines that connect all digits starting from "1" to "2" to "3" and so on (see figure 2.1 left). In part B, the participant has to connect letters and numbers alternating, that is "A" to "1" to "B" to "2" and so on (see figure 2.1 right). The experimenter stops the time and records mistakes/corrections made by the participant. Whereas both parts also measure the participant's ability of visual search and psychomotor speed, TMT-B is more sensitive to fluid intelligence like working memory, processing speed and inhibition abilities (Ashendorf et al., 2008; Czaja et al., 2006). Especially the TMT-B is therefore regarded as a good indicator for the degree of cognitive ageing. Administered after the TMT-B, it requires the participant to inhibit the previously learned rule (TMT-A: digit-digit-digit connections) and comply with the new rule (TMT-B: digit-letter-digit connections). Even though early research suggested that age did not influence TMT performance (Reitan & Wolfson, 1994), more recent work with more representative and larger samples provided normative data and showed the expected performance decline with age (Ashendorf et al., 2008; Ivnik, Malec, Smith, Tangalos & Petersen, 1996; Rasmusson, Zonderman, Kawas & Resnick, 1998).

Second, the **Digit Span task** has been shown to be a reliable and valid instrument for measuring working memory capacity (Wechsler, 1997) and is widely used in cognitive assessment and neuropsychological testing (Rabin et al., 2005). The participant is verbally presented with a sequence of digits (e.g. 1-5-8-2, usually one digit per second) that has to be repeated verbally in the correct order. Two variants can be applied: In Digit Span Forward, the sequence has to be repeated in the given order while in Digit Span Backward, the participant has to repeat the sequence in the reverse order. In general, the forward variant is less demanding compared to the

W	B	T	P	V	D	G	C	J
1	2	3	4	5	6	7	8	9

T	W	C	G	J	V	B	D	P	V	P	T	D	C	B

Figure 2.2: Coding scheme and first test line for the Digit Letter Substitution Test (van der Elst et al., 2006).

backward variant and performance declines in both variants with age (Grégoire & Van der Linden, 1997). After each completed trial, the number of digits that has to be recalled is increased by one and the test is stopped as soon as the participants cannot repeat two sequences with the same number of digits (Wechsler, 1997).

Third, the widely common **Digit Letter Substitution Test (LDST)** – as an adaptation of the earlier Digit Symbol Substitution Test (Wechsler, 1955) – relies on the two "over-learned" signs digits and letters and measures primarily information processing speed (van der Elst et al., 2006). Participants receive a letter-substitute for each digit (1-9; e.g., "1" is "W", "2" is "B", and so on; see figure 2.2) and have a limited time to substitute as many letters given in rows on a paper-sheet with their corresponding digits (the coding scheme is always visible for participants). The number of correct substitutions declines with age and normative data on large sample sizes is available (van der Elst et al., 2006). The advantage of LDST over the Digity *Symbol* Subsitution Test lies in the fact that participants only have to learn the association between two sets of symbols (digit and letter) but not the symbols itself.

It must be taken into account that older adults have been shown to be generally more cautious in standardised cognitive tests and focus on accuracy instead of speed (Brébion, 2001) and that sensorimotor speed has an impact on the performance in cognitive tests (Rabin et al., 2005). However, it is well established that the here presented measurements are valid and reliable tools for assessing the cognitive ability of people from different age groups. Especially the TMT (parts A and B) is pervasive in neuropsychological testing, can be easily applied in HCI research and covers core predictors of fluid intelligence. Besides this excerpt, some other cognitive tests are available that could be (and have been) also used for quantifying the effect of age-dependent cognitive decline. Most prominently, the reader is referred to the Wechsler scales as the most pervasive assessment tool for fluid and crystalline intelligence (Wechsler, 1997) or the publication of Czaja et al. (2006).

2.2.3 Impact of cognitive ageing on HCI

Older adults are likely to face cognitive ageing, but how does that influence their interaction with technology? Several studies show that learning new technologies and interaction performance can be severely affected by cognitive abilities (Barnard et al., 2013; Chevalier et al., 2012; Czaja et al., 2006; Hanson, 2011; Wagner, Hassanein & Head, 2014; Zhang, Grenhart, McLaughlin & Allaire, 2017) and Czaja and Lee (2007) underline the importance of cognitive abilities to cope with innovative and unfamiliar forms of HCI. The following section, therefore, provides some examples shaping the challenges resulting from cognitive ageing that will be addressed in this work. Most importantly, it is argued that cognitive abilities are crucial for learning the interaction with unfamiliar technology and cognitive decline is opposed to effective and efficient interaction.

Since chronological age and cognitive abilities correlate (Salthouse, 2010) studies on ageing and HCI often use only chronological age as a proxy for cognitive abilities (Dickinson et al., 2007; Vines et al., 2015). As described above, direct measurements like cognitive abilities are more appropriate when conducting comparative studies between younger and older adults. In the following, recent examples of research that directly investigated the impact of specific constructs of cognitive abilities on HCI are summarised.

For example, Wagner et al. (2014) found that cognitive decline in the form of declining spatial abilities (measured via the performance in recreating the spatial hierarchy of a website) mediated the effect of age on performance measures when navigating through websites. Even though they solely focused on the spatial abilities and forming of spatial mental models, they could provide evidence for the influence of cognitive abilities on performance measures.

Crabb and Hanson (2014) as well as Crabb (2013) also stressed the point of directly measuring cognitive abilities as well as prior technological knowledge to differentiate the effects of age on HCI further. They applied several tests from the test-battery of the seminal CREATE-project (see section 2.3) and measured fluid intelligence, processing speed, short-term memory and long-term memory (as well as prior exposure to the Internet). They found that cognitive abilities were a better predictor than age for differentiating between users and their performance in an online search task.

Chevalier, Dommes and Marquié (2015) also underlined the importance of cognitive abilities for HCI. They investigated how cognitive abilities influence the performance in search tasks (using a search engine) of differing complexity and found that with increasing age, the time to complete search tasks did not increase (efficiency) but the number of correct search results decreased (effectiveness). The authors suggested that even though older adults spent the same time on the task, they tended to perform an exhaustive evaluation of the user interface and the content of the search request, a phenomenon which has been reported earlier (Aula, Khan & Guan, 2010). Cognitive abilities (e.g. working memory measured via the TMT-B, see figure 2.1) was seen as an important factor inhibiting the successful completion of user tasks.

Other work refers to the impact of age-related changes in cognitive abilities on HCI but without directly measuring cognitive abilities or performance data (Kuerbis, Mulliken, Muench, Moore & Gardner, 2017; Page, 2014; Zhang et al., 2017). For example, on the user interface level, Chevalier et al. (2012) compared the two effects of a) ageing and b) conformity with website ergonomics guidelines on the user performance. Younger and older participants completed different use cases either on a “user-friendly” or “non-user friendly” website. Besides standard usability-criteria (time, number of correct answers), they also tested the mental model built during the experiment through recall of specific interface-elements. As expected, both age and non-conformity to guidelines hurt performance (e.g., visual search). No interaction between the websites’ conformance with guidelines and age was found for task times and number of steps. However, only non-conformity with guidelines but not age per se affected the building of a correct mental model, showing that the tendency of older adults to deeply understand things (like an interface) before using it to solve tasks (Aula et al., 2010). Even though age differences are smaller when the user interface is less complex (Ziefle & Bay, 2005), studies from different domains such as mobile health-care (Kuerbis et al., 2017), kitchen devices as microwaves (Blackler et al., 2010), touchscreen-based applications (Page, 2014), or reminder systems (Razak & Razak, 2013), support the fact that older adults are at disadvantage when it comes to learning and interacting with new interfaces and cognitive abilities are usually seen as one of the core factors mediating the age-effect on effectiveness and efficiency in interaction.

2.2.4 Summary: Cognitive ageing

Cognitive decline is a major problem for HCI and has received increased attention in the recent past. Older adults face a number of challenges when interacting with new technologies. Different theories try to explain lower cognitive performance for older adults. Most importantly, changes in working memory, processing speed and the ability to inhibit irrelevant information were identified as major markers of cognitive abilities. Important measurements have been described that can be regarded as standard procedures in cognitive assessment. Still, cognitive abilities alone are not sufficient for explaining the poorer performance with technology of older adults. The next section introduces a second significant difference between younger and older adults that affect their interaction with technology.

2.3 Ageing and Prior Technological Knowledge

Besides declining cognitive abilities, older adults statistically differ in another important aspect from younger adults raising further challenges for HCI: compared to younger adults, they are less likely to have extensive prior exposure to and thus knowledge about modern technologies (Barnard et al., 2013; Czaja et al., 2006). This section, therefore, provides an overview over the most critical factors contributing to technology use and how prior technological exposure (though not always resulting in prior technological *knowledge*) can influence effectiveness and efficiency of HCI.

2.3.1 Age-effects on technology adoption and acceptance

It is estimated that each human being will use 20,000 different objects throughout the lifespan (Norman, 2013). Most of them are highly specialised and have to be learned by the human being and many of them being based on today's innovative technology. Of course, older adults use modern technologies, too (Vorrink et al., 2017). O'Brien (2010) reported between 150 and 300 interactions with technologies within ten days for older adults between 65 and 75 years. Most prominently, the Internet has become pervasive and adults between 50 and 70 years use it on a regular basis (Damodaran, Olphert & Sandhu, 2014; A. Smith, 2014; Vorrink et al., 2017). Older adults are the fastest growing internet population (A. Smith, 2014). In 2015, 90% of older adults between 50 and 64 years (mainly stemming from the generation of baby boomers) possessed a cellphone compared to 78% of adults over 65 years (Kuerbis et al., 2017). Smartphones were less common, but still, 58% (50 to 64 years) and 30% (65+) reported to use a smartphone. Interestingly, both age groups seem to be similar in their use of tablets (37% and 32%), E-Mailing (91% and 87%) and SMS/Texting (both 92%).

However, the stereotype exists that older adults are less likely to use technology (Olson, O'Brien, Rogers & Charness, 2011) and older adults might even be affected by a self-fulfilling prophecy ("I am too old to learn this new technology") which in fact negatively influences their performance due to adverse self-efficacy regarding new technologies (Barnard et al., 2013; Chen & Chan, 2013). Even though this is not true for each individual (Barnard et al., 2013; Mitzner et al., 2010), older adults, as defined by chronological age, in fact statistically use fewer different technologies compared to younger adults (Czaja et al., 2006; A. Smith, 2014; Vorrink et al., 2017) and they use not only fewer technologies but these also with a lower frequency (Mitzner et al., 2010).

One reason for this is linked to cognitive abilities as described in section 2.2. Using a substantial and representative sample of 1204 participants between 18 and 91 years, Czaja et al. (2006) found that technology use decreased with age. Participants completed a number of cognitive tests as well as questionnaires on prior exposure with technology and attitudes towards computers. The authors report that younger adults used more technologies compared to older adults, but also decompose this age-effect into several mediator variables. Both higher fluid and crystalline intelligence were significant predictors (fluid more than crystalline) of more technology use. Also, younger adults scored higher in measurements for fluid intelligence and older adults higher in measurements for crystalline intelligence. Additionally, higher computer anxiety was negatively and higher computer self-efficacy positively correlated with technology use. Since older adults reported higher computer anxiety and lower self-efficacy than younger adults, these two factors might explain the lower technology use of older adults. They concluded that cognitive decline (here, the decrease of performance in tests of fluid intelligence) has a direct negative impact on technology adoption which is seen as extremely important for an independent lifestyle.

However, other factors are critical why “older adults are slower to adapt to newer technologies” (Vorrink et al., 2017, p.1). Based on a long tradition on the topic of technology adoption, several explanations have been put forward for the fact that older adults use less technology compared to younger adults (Chen & Chan, 2014; Davis, 1989; Venkatesh et al., 2003). Most prominently, the Technology Acceptance Model or TAM (Davis, 1989) serves as a robust framework for explaining the differences in technology adoption between individuals or user groups. Even though further developed into the TAM2 (Venkatesh & Davis, 2000) and the Senior Technology Acceptance Model or STAM (Chen & Chan, 2014), it basically proposes two major factors that have to come together to increase the probability that a target group will use a technology: perceived usefulness and perceived ease of use. On the one hand, perceived usefulness of the new technology is essential for motivating potential users. The technology must match a need and solve a real problem, otherwise, the added value remains questionable for the target group. On the other hand, the perceived ease of use facilitates the adoption of a new product. Perceived ease of use refers to the hassles a new user is confronted with and a low perceived ease of use decreases the probability that the target group will use the technology again. Perceived ease of use can also influence the perceived usefulness: when the effort to use a new technology is too high, it will be less useful than when it is low. A variety of external circumstances can have an impact on these two main factors (e.g. social influence, costs, prior exposure). But if a new technology is perceived as useful and easy to use, the TAM predicts that a positive attitude and an intention to use a new technology will be formed. Finally, this will lead to the real usage of technology.

Based on the TAM, a more comprehensive framework for describing and predicting technology use has been proposed, which can be applied to as well younger and older adults: the Unified Theory on the Acceptance and Use of Technology or UTAUT (Venkatesh et al., 2003). Similarly to the TAM or STAM, the perceived usefulness (here called expected performance) and the perceived ease of use (here called expected effort) are essential factors for forming an intention to use technology (see figure 2.3). In UTAUT, an additional third factor is highlighted: social influence. Even though the UTAUT includes more factors and is more comprehensive, it has been criticised as being less parsimonious than the previous TAM and its extensions (Van Raaij & Schepers, 2008).

In its essence, three important factors can be distilled for the intention to use a technology and finally the use and adoption of technology: performance expectancy, effort expectancy of use and social influence. There is a growing need for older adults to use technology in order to participate in social and work life and especially these factors can provide a proper basis for understanding existing facilitators and barriers to technology adoption amongst this age group.

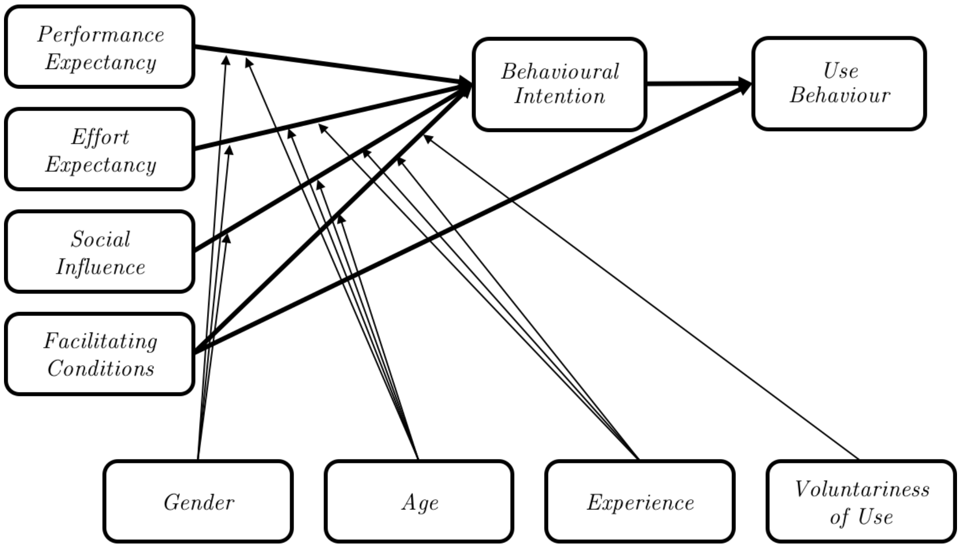


Figure 2.3: Unified Theory of Acceptance and Use of Technology, after Venkatesh et al. (2003).

Social influence. For example, communication with peer groups or family shifts from face-to-face meeting towards digital devices and older adults are socially urged to adopt the new forms of communications if they want to remain connected over large distances with their children (Hardill & Olphert, 2012; Hope et al., 2014), which is a constant concern for many older adults (Bailey & Sheehan, 2009). Elliot and Maier (2014) even report that the use of communication devices positively correlates with over 65 years old adults' well-being and absence of diseases, a fact that the authors attribute to a stronger connectedness to their social environment. Regarding smartphones, older adults often see them as "liberating" compared to younger adults who describe smartphones sometimes as "leashes" that bind them (Center, 2015). In extreme cases, some older adults even welcome digital communication because it "keeps their families out of their house and business" (Bailey & Sheehan, 2009). Also, technological advances such as in the health-care sector (e.g., monitoring equipment like a wireless heart monitor) enable older adults to stay longer independent from dedicated caregivers safely. These devices are often required to be operated by older adults and promoted in their peer group (Kang et al., 2010). Finally, digitalisation of current workspaces confronts also older adults with novel technologies and most baby boomers over 50 must use a computer in their daily work (LeRouge, Van Slyke, Seale & Wright, 2014). Social influence can support or inhibit technology use of older adults. Stigmatising older adults due to age-specific technologies – such as a highly distinguishable "Senior Smartphone" – could also decrease the probability of its adoption by the target group.

Performance Expectancy. Besides the external social influence, older adults also use technology intrinsically motivated. In fact, older adults perceive novel technologies mostly as something positive and valuable (Mitzner et al., 2010). Older adults highlight the importance of technology's usefulness and ease of use and seem to build their attitudes and intentions to use a technology directly on perceived usefulness (Barnard et al., 2013; C. Lee & Coughlin, 2015; Lüders & Brandtzæg, 2017; Melenhorst, Rogers & Bouwhuis, 2006). Also, technology use by older adults differs between domains. Technology that has been available on the market for a longer time (e.g. telephone, microwave) also has a high probability of being also used by older adults (Olson et al., 2011; Vorrink et al., 2017). In sum, when older adults can adopt technology or digital activities as long as they perceive them as useful and performing well for their needs (Vroman, Arthanat & Lysack, 2015; Wandke et al., 2012).

Effort Expectancy. The third important factor for technology adoption focuses on the expected effort to use the technology or the perceived ease of use. Older adults have been found to prefer learning to use tablet applications over learning to use desktop computer applications given especially their ease of use (Jayroe & Wolfram, 2012). Older adults seem to be willing to learn and use also innovative technologies, but favour ease of use over the efficiency of use (Barnard et al., 2013). However, older adults tend to approach new technology more passively compared to younger adults, constrain their actions due to expected failure and restrict themselves in what they attempt in the first place (Hawthorn, 2007). When confronted with obstacles in HCI, W. A. Rogers, Meyer, Walker and Fisk (1998) report that older adults ceased their activity (i.e., the task is not done or delegated to another person) in 51% of all cases, while only the minority tried again until they succeeded (22%), sought compensations by assistive technologies (19%), or tried to improve themselves by training on the long-term (8%). Ideally, older adults interact successfully at the first encounter instead of trying it only once and then pass it on to their children to help them. In general, older adults are slightly passive in HCI, trial and error are not their strategy when using a novel technology and possess an HCI-related "fear of the unknown" (Hawthorn, 2007, p.336).

A low effort expectancy is therefore crucial for the first interactions with new technology to prevent a demotivation of further technology use. However, as it has been shown in sections 2.1 and 2.2, coping with new forms of HCI is more demanding for older adults due to statistically declining sensorimotoric and cognitive abilities.

Importantly, effort expectancy can be linked to the user's prior technological exposure and knowledge (for the difference, see 2.3.2). For this, the concept of intuitive use – i.e. the application of prior knowledge using only a minimum of cognitive resources leading to an effective and satisfying HCI (Blackler & Popovic, 2015; Naumann et al., 2007) – is of high relevance.

Older adults have been shown to be less familiar with modern technology and thus can draw on less prior technological knowledge compared to younger adults (Czaja et al., 2006; Vorrink et al., 2017), which decreases the chance of intuitive use of new HCI for older adults. Without the experience of intuitive use, a low perceived ease of use will result in a decreased probability of user adoption in the long term. A decreased user adoption will again inhibit further technological exposure and the forming of prior technological knowledge: a vicious circle.

Within this framework, this thesis focuses primarily on the dimension of perceived ease of use and effort expectancy, since it is critical to successfully interact with a new interface in the first place to have at least the chance to cope with it later on. As a consequence, a high learnability has been proposed to be important especially for older adults when it comes to HCI (Barnard et al., 2013; Fisk et al., 2009). Since prior knowledge has been shown to be essential for intuitive use, older adults should be at a disadvantage concerning intuitive use when being confronted with new technology, which should lead to lower ratings of perceived ease of use. But what is prior knowledge? Because intuitive use is that tightly linked to prior knowledge, the following sections give an overview of different levels of prior exposure and knowledge that provide the basis for intuitive use in the context of age-inclusive HCI.

2.3.2 Prior exposure versus prior knowledge

Prior knowledge, prior experience and prior exposure have been used often interchangeably and referring to similar meanings (Hurtienne et al., 2013; O'Brien, 2010; Reddy et al., 2014; Sanchiz, Chevalier, Fu & Amadieu, 2017), but are in fact different things. Based on the definitions of Hurtienne et al. (2013) and Reddy et al. (2014), in this work, prior *knowledge* is understood as something that can (but does not have to) result from previous *exposure* with technology, but also with general things like physical objects, plants or animals. Prior knowledge is thus a crystallised and structured form of previous exposure that can – regarding intuitive use – be applied and extended to new environments and tasks (Blackler & Popovic, 2015; Naumann et al., 2007). However, prior exposure with technology alone is neither sufficient nor necessary for obtaining prior knowledge that is relevant for designing intuitive interactions.

Older adults often do not show high competence in their everyday technology (knowledge) even when their exposure to it is high (Lawry, 2012). Younger adults are more likely to learn a new technology when exposed to it (see also the concept of technology generations in section 2.3.3), which might be explained by younger adults' more cognitive abilities that are necessary for turning exposure into knowledge (Hanson, 2011; Langdon, Lewis & Clarkson, 2010; Lewis, Langdon & Clarkson, 2008). Zhang et al. (2017) also emphasised the link between cognitive abilities and computer competence.

Hurtienne et al. (2013) thus distinguish clearly between a) exposure to technology and b) technology-related competence. In their understanding, measuring exposure to technology provides insights about duration, intensity, and diversity of technology use while the dimension of competence directly addresses the more critical question of acquired skills and knowledge. For an intuitive use of new technologies, the level of competence is thus more important compared to exposure. Several instruments have been suggested to measure prior exposure and knowledge that is relevant for HCI such as, for example, the Computer Literacy Scale (Sengpiel & Dittberner, 2008), the

Computer Proficiency Questionnaire (Boot et al., 2013), the competence dimension of the Questionnaire on Technology Affinity (Karrer et al., 2009) or more recently the Mobile Device Proficiency Questionnaire (Roque & Boot, 2016). Importantly, prior technological knowledge can only predict a certain degree of intuitive use and prior exposure to technology is not always necessary for prior knowledge that can be relevant for intuitive HCI.

2.3.3 Forms of prior knowledge

Following Naumann et al. (2007) and their knowledge continuum (see 2.4), prior knowledge that serves as the basis for intuitive use can be differentiated into several levels. On the highest level (expertise) we find knowledge about complex interfaces such as the cockpit of an aircraft or also standard gestures on a tablet. One level below (culture), shared norms and standard convention such as the interpretation of the colours of traffic lights are usually known by a larger group of people and learned by the majority of people in one population. The third level (sensorimotoric) is regarded as learned automatically by almost all people throughout their lives and covers physical facts like gravity or reoccurring, elementary patterns in the world like object permanence. The fourth and lowest level (innate) does not have to be learned and reflexes are one form of prior knowledge that falls into this category. Even though the borders between the four categories can be seen as not entirely separating and not always clear (Naumann et al., 2007), this framework allows for a rough segmentation of different forms of prior knowledge.

The four levels require different forms of experiences, limiting the group of people that has access to each of the four levels of prior knowledge. Prior knowledge from the innate level requires no external experience and, therefore, should be found in every human being. Prior knowledge from the sensorimotor level is formed from repeated encounters with physical concepts in the environment (e.g. gravity, containers) which means that almost every individual, regardless of culture or age, should be able to draw on the same sensorimotor knowledge as soon as the individual did interact with the physical environment. Lower levels of prior knowledge ground on understanding fundamental characteristics of our environment that humans encounter over and over. Due to a high frequency of the encoding and retrieval (e.g. concept of gravity: a stone falls down, a bottle falls down, a leaf falls down), this form of prior knowledge can be regarded as automatic in retrieval.

On the contrary, the cultural level is accessible only to individuals from the specific cultural background and the level of expertise only to few people that invested time to learn operating a specific device or artefact. Because the experiences necessary for forming this form of prior knowledge are not as ubiquitous as physical laws, encoding and retrieval of this form of prior knowledge is less frequent and limited to a smaller group of people. In sum, lower levels of prior knowledge are more ubiqitous than higher levels of prior knowledge. However, the pervasiveness of lower levels of prior

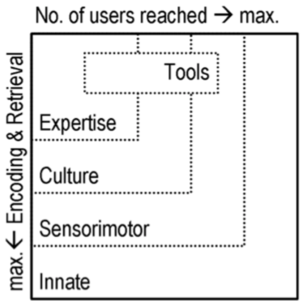


Figure 2.4: Knowledge Continuum (Hurtienne, 2011; Naumann et al., 2007).

knowledge is very positive: when almost all people have access to the same sensorimotor and innate prior knowledge, exploiting this form of prior knowledge in HCI promises interfaces that include a wide range of people that still can differ in higher levels of prior knowledge (e.g. technological prior knowledge).

The continuum of knowledge in intuitive interaction can serve as a framework for describing different levels of prior knowledge that can be applied to – also new and innovative – technology. The value of different forms of prior knowledge can be estimated by using the so-called domain transfer distance (Diefenbach & Ullrich, 2015). For example, prior knowledge from a ticket vending machine (TVM) in Madrid can be applied when the same TVM in Madrid (but at another station) has to be used. Here, the domain transfer distance is near to zero. Also, some prior knowledge about a TVM in Berlin can also be applied to the TVM in Madrid, but the domain transfer distance is to be expected as higher compared to in the first case. Prior knowledge about touchscreens in general (e.g. obtained from an Automatic Teller Machine) can also be applied to the TVM in Madrid. Still, the domain transfer distance is considerable providing limited value concerning intuitive use. Finally, paying attention to blinking lights (e.g., to the output box of the ticket vending machine) can also be seen as a fundamental form of prior knowledge that is relevant for operating the ticket vending machine (thus, even ambient has been shown to allow for an intuitive and unconscious basis of HCI building on the most basic level of prior knowledge; see Tscharn, Ly-Tung, Löffler & Hurtienne, 2016; **?**).

Taking into account differences between younger and older adults, another classification can be made that frames the knowledge continuum differently. **Technology-dependent prior knowledge** on the one hand can be often readily applied to mostly technology-based HCI. Hurtienne et al. (2013) could show that especially technology-dependent prior knowledge (i.e. competence) that is not too specific to the technology of interest but also not too far away from it is the best predictor for task performance and success. The value of prior knowledge is closely linked to facilitating factors such as the recency and frequency of access (Reason, 1990). Also, the importance of prior knowledge for HCI has also been classified by Docampo Rama, Ridder and Bouma (2001) in terms of *when* the prior knowledge was formed. They suggest that the most fundamental technology-dependent prior knowledge is obtained in the human's formative period between 10 and 25 years. The form of technology a human is exposed to in this period shapes their prior knowledge for the rest of their lives. For example, in the pre-electrical era (people born before 1960, Docampo Rama et al., 2001), where the input was directly linked to the output (lever, wheel, gears), a fundamentally different form of technology-dependent prior knowledge was obtained than in the electrical era (people born after 1960), where many actions became abstract (one button in a graphical user interface can be linked to any function). As a consequence, younger and older adults have access to entirely different forms of technology-dependent prior knowledge and today's younger adults will again differ from the next generation in 50 years. Thus, the problem of a lack of prior technological knowledge will not vanish in the future. Instead, the prior technological knowledge of today's younger adults (the "digital natives") will be of little help when they have become older adults themselves. In a diary study, O'Brien (2010) investigated how often older adults reported prior knowledge as the reason for not being able to operate a technology. Besides expected severe constraints due to perceptual or motoric impairments, they found that cognitive abilities were not as crucial for the intuitive use of technology compared to prior knowledge about similar technologies.

Since technology-dependent prior knowledge is so crucial for successful HCI, research has also shed light on means on how to increase technology-dependent prior knowledge by training. Struve and Wandke (2009) showed on the example of ticket vending machines that older adults profit from video-based learning (especially when the training includes guided error-making) and Struve (2010) provides guidelines for instruction design for technologically less-experienced older adults. Bruder, Blessing and Wandke (2014) further investigated adaptive learning for older adults and found that a dynamic adaption of the user interface complexity (but not the adaption of advices) during during the learning phase had a positive impact on the success of learning how to use a mobile phone. In sum, technology-dependent prior knowledge is often missing and severely inhibits HCI, but can be provided by means of training. However, another way to counteract this lacking prior knowledge is to take into account technology-independent prior knowledge.

Technology-independent prior knowledge on the other hand draws on lower levels of the knowledge continuum such as the sensorimotor level. Younger adults have been found to possess more technology-dependent prior knowledge than older adults. Thus, one way to provide both younger and older adults with intuitive HCI might be found in grounding this HCI on lower levels of prior knowledge that is stemming from technology-*in*dependent prior knowledge. In other words, people of different age-groups should have the same sensomotoric knowledge and exploiting this lower level of the knowledge continuum for interaction design should lead to solutions that are more inclusive and independent from cognitive abilities and prior technological knowledge. Even though older adults do not have the technology-dependent prior knowledge to interact with new technologies intuitively and perceive them as easy to use, technology-independent prior knowledge might be the source of prior knowledge that is accessible for older adults and younger adults at the same time.

Research shows that technology-dependent prior knowledge can predict intuitive use with technology (Blackler et al., 2012; Langdon et al., 2010; Lewis et al., 2008). Since younger and older adults differ in their degree of technology-dependent prior knowledge, this form of prior knowledge cannot serve as a basis for age-inclusive design. However, little work has been done to achieve age-inclusive interaction design by grounding interfaces on technology-independent prior knowledge. But there certainly are areas from which technology-independent prior knowledge can be borrowed. Besides the knowledge-continuum, a set of technology-independent knowledge areas has been put forward by Jacob et al. (2008). They argue that, in HCI, users could exploit prior knowledge about a) naive physics, b) body awareness and skills, c) environment and skills, and d) social awareness and skills. Based on their framework of so-called *Reality-Based Interaction*, modern forms of HCI can be understood in terms of their applied technology-independent prior knowledge. Still, this approach might be too broad and not specifc enough to guide interaction designers on the interface level. As chapter 4 will describe in more detail, the concept of image schemas and image-schematic metaphors can provide an operationalisation of this technology-independent prior knowledge.

M. Johnson (1987) argued that through everyday experiences in the physical world all humans (i.e. younger and older adults) acquire a mental set of basic concepts, which we use to make sense of the world that surrounds us. The elemental building blocks of this set are called image schemas and help us to structure our perception and mental representation of the world. For example, we repeatedly perceive objects as being in- or outside of containers, up in the air or down on the ground, permanently linked or having temporary contact with other objects. These perceptions form image schemas (e.g. IN-OUT, UP-DOWN, LINKAGE, CONTACT) that shape our language, cognition, and behaviour. Following Johnson, image schemas are not constrained to understanding only our physical environment. Moreover, we also use image schemas when we think and speak about abstract concepts that do not have physical equivalents per se. For example, we talk about *rising* inflation, even though inflation cannot be perceived in the physical world as something moving from up to down or vice versa. These associations between abstract concepts and image schemas can be exploited in interaction design as a very basic form of technology-independent prior knowledge (Hurtienne, 2011).

2.3.4 Summary: technology use and prior knowledge

Older adults have been shown to adopt new technology at a slower pace and to have less prior exposure to technology. As a consequence, older adults are likely also to have less technology-dependent prior *knowledge* compared to younger adults. Since prior knowledge is a core pre-requisite for intuitive use, the perceived ease of using new technologies – a significant predictor of adopting new technology – is often reported as low and especially older adults are in danger of being further excluded from new technology. Therefore, other forms of prior knowledge that are not related to technological exposure should be exploited.

2.4 Designing for Older Adults

Regarding age-related changes, many guidelines have been proposed to build interfaces that enable older adults to interact with. Often, these remain at the level of accessibility guidelines and focus on the mere perceivability of the interface elements. For example, colour and contrast perception declines with age and older adults experience difficulties when differentiating between the colours blue and yellow (Haegerstrom-Portnoy, Schneck, Lott, Hewlett & Brabyn, 2014). Findings like these allow for deriving design recommendations like “Avoid using blue elements next to yellow elements in the user interface”.

Documented impairments have been translated into large sets of guidelines for interaction designers (Brajnik et al., 2011; Fisk et al., 2009; Nurgalieva et al., 2017). The WCAG (Web Content Accessibility Guidelines), which is the currently most pervasive standard of recommendations for the accessible design of web-content, also embraces many of these guidelines (Caldwell, Cooper, Reid & Vanderheiden, 2008). Also, many guidelines and adaptations for participatory design for older adults are available that focus on HCI for older adults (Fisk et al., 2009; Reddy et al., 2014; Vines, Blythe, Dunphy & Monk, 2011; Zaphiris, Sustar & Pfeil, 2008). Since early days of HCI, guidelines have thus been reviewed and summarised. For example, already in 1986, S. L. Smith and Mosier (1986) provided a collection of guidelines to facilitate the general design process of user interfaces not specific to older adults, but age-specific large sets of guidelines are available as well (Fisk et al., 2009; Mi, Cavuoto, Benson, Smith-Jackson & Nussbaum, 2014; Nurgalieva et al., 2017).

Guidelines on the focus of this work – age-related differences in cognitive abilities and prior technological knowledge – are not as comprehensive as in other domains such as vision or auditory accessibility. For addressing challenges in HCI, based on cognitive impairments of older adults, Nurgalieva et al. (2017) summarise 74 guidelines that have been reported in the literature. Interestingly, even though these guidelines primarily focus on older adults with mild cognitive impairments, they also address other user groups such as people with general learning impairments. Reported guidelines range from “Include built-in instructions and an example video, to support independent learning and use.” (Martin, Laird, Hwang & Salis, 2013) to “Allow tasks to be accomplished serially, don’t force them to be done at the same time requiring cognitive switching.” (Carmien & Manzanares, 2014) and “Make feedback readable as older adults tend to have difficulty in interpreting the meanings of symbolic representations” (Kobayashi et al., 2011).

Other guidelines address the prior knowledge of older adults: ”Consider (and evaluate early) cultural issues in the design of icons, particularly given users’ age and strong attachment to a specific culture.” (Neves et al., 2015) and “Avoid scrolling and requiring double click. It is not intuitive, it should be learned.” (Carmien & Manzanares, 2014). Especially for age-related differences in cognitive performance and prior knowledge, Reddy (2012) investigated in two studies interface characteristics and their impact on age-inclusiveness. Specifically, he compared hierarchical menu-structures to flat one and found that differences between age-groups were smaller for flat ones than for hierarchical ones. Also, he found that differences between age-groups were smaller for interfaces that used icons without descriptive text, but that differences between younger and older adults increased when icons were accompanied by a descriptive text that had been intended actually to decrease age-differences.

It is important to note that complying with accessibility requirements ensures only that information is accessible and that missing “communication channels” (e.g., vision) between a digital device and the user are compensated via other channels (e.g., auditory). Ben Shneiderman, a pioneer in universal usability, thus states that besides accessibility, another factor is vital for future computational devices: “inclusiveness, what I call universal usability, enabling all citizens to succeed in using information and communication technologies to support their tasks” (Shneiderman, 2003, p. 14). Importantly, he does not state that people should be enabled to use but to be enabled to succeed in using technology. Even though accessibility is a prerequisite for universal usability, the two terms must be distinguished. However, how can universal usability and inclusiveness be designed? The answer is called inclusive design. Inclusive Design is defined as the ‘design of mainstream products and/or services that are accessible to, and usable by, people with the widest range of abilities within the widest range of situations without the need for special adaptation or design” (Keates, 2004, p. 333). General frameworks or design recommendations for older adults are still often limited to accessibility issues or focus only on older adults and ignore threats of stigmatisation by senior-specific interfaces. However, inclusive design is more than applying design guidelines that address specific user populations. Instead, inclusive design aims at universal usability and suggests including of a variety of users in the development process to increase the probability of covering the maximum of target groups with a product or user interface.

Interaction designers are prone to primarily designing for themselves (Hawthorn, 2007; Vines et al., 2015) and requirements by older adults are likely to be ignored during product development. General guidelines, user research and User-Centred Design (see also chapter 3) can provide a remedy for this. Still, HCI research and interaction design are often separated between disciplines (e.g. psychology and design) and insights from user research are in danger of not being appropriately communicated to the interaction designer. For example, Zimmerman, Forlizzi and Evenson (2007) report on an inquiry in the HCI-community that “many HCI researchers commonly view design as providing surface structure or decoration” (p. 1). During their data collection, including a CHI-workshop, they found a wide gap between two principal groups: researchers on the one hand, who provide – ideally valid – generalisable requirements based on experiments and guidelines, and practitioners and interaction designers on the other hand, who do not adopt these requirements and guidelines which massively limits the impact of them. Even though teams become more and more interdisciplinary, this gap has not been significantly closed in the last years (Zimmerman & Forlizzi, 2014).

Holtzblatt and Beyer (2017), focusing on User-Centred Design, also urge project teams aiming at age-inclusive design not to underestimate the gap between user research and design since especially visual designers are often not trained well enough in proper methodology to draw full potential of these methods. Guidelines are relevant, but without providing a proper and understandable framework for applying them, much of their potential is given away. Moreover, general guidelines might even be seen to interfere with innovation and newness since they narrow down the design space of possible solutions. By providing recommendations for accessible solutions, the designer is restricted in creativity and must follow – in the worst case – a set of standardised requirements that try to realise accessibility in technology.

In sum, guidelines are essential contributions to age-inclusive interaction design, even though they primarily focus on adaptations of interfaces to the needs of older adults and not on designing interfaces that decrease differences between age-groups. The approach of User-Centred Design has been proposed to achieve age-inclusive design, (i.e., one solution for younger and older adults). Even though User-Centred Design is currently the major methodology in the context of age-inclusive interaction design (Langdon et al., 2014) and early user involvement has been shown to be the key to inclusive design in general. However, especially for differences in cognitive abilities and prior technological knowledge, no framework is available that can facilitate the interaction design process. However, Nurgalieva et al. (2017) underline the fact that guidelines and frameworks that deal with accessibility are available; but for a real-world impact, it is essential to make these guidelines and frameworks accessible to researchers and practitioners as well or provide methods that allow for a smooth integration into existing work procedures.

2.5 Summary

Older adults are a very diverse user group and face many challenges compared to younger adults. Besides sensorimotoric impairments, two factors affect the interaction with technology: cognitive ageing and decreased technology-dependent prior knowledge. Cognitive ageing is well reported in the literature and has been linked to less effective and efficient interaction with technology. Also, older adults use modern technology less frequently than younger adults. This fact has been explained by various factors including a low perceived usefulness or perceived ease of use regarding new technologies. Less technology use and exposure imply less technology-dependent prior knowledge that can be re-used for interacting with new technology. Additionally, exposure to technology has been shown to lead to less prior knowledge for older adults compared to younger adults, possibly due to cognitive constraints turning exposure in applicable knowledge. Taken together, for age-inclusive interaction design, the two factors cognitive ability and prior knowledge must be addressed in HCI.

Chapter 3

Designing for Innovation

"Companies that do not innovate die" (Chesbrough, 2006, p. xvii): whether in services, products, technologies, economic systems, or organisational structures, innovation is a positively connoted buzzword, and the field of Human-Computer Interaction (HCI) is no exception (Hassenzahl, 2004; Norman & Verganti, 2014). Even though they use other terms, people have strived for being "innovative" for a long time (Lorenzi, Mantel & Riley, 1912; E. M. Rogers, 1995). Innovation is tightly linked to newness and novelty (Radford & Bloch, 2011). Innovative approaches and solutions are at the core of interaction design (Biskjaer et al., 2010; Dalsgaard, 2014; Norman & Verganti, 2014). Older adults are frequently at a disadvantage in coping with innovative interaction design (see chapter 2). Innovation and age-inclusive interaction design is, therefore, a balancing act that cannot easily be solved.

This chapter summarises research on the construct of innovation (section 3.1) and *perceived* innovation and newness (section 3.2). Also, it discusses inter-individual differences iregarding the acceptance of innovations (sections 3.3). After making explicit the design problem (section 3.4), methods aiming at innovative interaction design on different levels are discussed (section 3.5). Finally, this chapter connects the concept of perceived innovation back to age-related changes relevant for HCI (section 3.6).

3.1 What is "Innovative"?

Adams, Bessant and Phelps (2006) laments on the fact that "the term 'innovation' is notoriously ambiguous and lacks either a single definition or measure" (p. 22). In fact, definitions of innovation differ both within as well as between domains and disciplines such as economic management, marketing, product design, rendering innovation a vague concept despite attempts to find one universal definition (Baregheh, Rowley & Sambrook, 2009).

Various disciplines coined the construct of innovation. As a consequence, several frameworks are available that try to both differentiate and condense existing descriptions of the construct of innovation. One often shared common ground is a clear focus of innovation on newness. In early days, the economist Joseph Schumpeter (Schumpeter, 1939), the anthropologist Homer Barnett (Barnett, 1953) as well as later on the sociologist Everett Rogers (E. M. Rogers, 1995) emphasised that novelty and newness are central for innovation (in their respective domain). Most universally, Barnett (1953) stated that innovation is "any thought, behaviour or thing that is new because it is qualitatively different from existing forms and is the basis of cultural change". E. M. Rogers (1995) saw innovation as "[...] an idea, practice, or object perceived as new by an individual or other unit of adoption" (p. 12).

Slappendel (1996) also used the concept of newness to distinguish innovation from mere change. Building on the differences between change and innovation opens an important dimension: incremental vs radical innovation. Incremental innovations are based on "improvements within a given frame of solutions (i.e., doing better what we already do)" while radical innovations are a "change of frame (i.e., doing what we did not do before)" (Norman & Verganti, 2014, p. 82). To become a radical innovation, the newness of a product is not sufficient according to Dahlin and Behrens (2005). They propose three criteria that must be met for a radical innovation: it must be a) novel, b) unique and c) adopted. In other words, a radical change – defined in a) and b) – must also be adopted by people and consequently penetrate the market – c) – to become a radical innovation. Radical innovation is rare, but incremental innovation is common. Interestingly, Norman and Verganti (2014) argue that radical innovation can tap entirely new opportunities for HCI (which are not reachable with incremental innovation), but only incremental innovation can draw full potential of them.

Besides newness, market penetration can be seen as the second important factor for qualifying as an innovation (Kanagal, 2015; Kelley et al., 2013; E. M. Rogers, 1995). E. M. Rogers (1995) introduced the Innovation Diffusion Theory (IDT), a framework for describing the process of market penetration of innovations as a social process influencing the individual's decision making of adoption or not. It distinguishes five stages of diffusion:

1. knowledge about an innovation
2. persuasion that the innovation is positive or negative
3. decision in favour or against the innovation
4. implementation of the innovation and using it
5. confirmation that the decision was right (and using it further on) or wrong (and reverse the decision to use it)

Individuals differ in their a priori probability of their decision making in favour of the innovation (e.g. being surrounded by a stimulating peer-group that facilitates step 1: "knowledge about an innovation" or inter-individual differences in attitudes

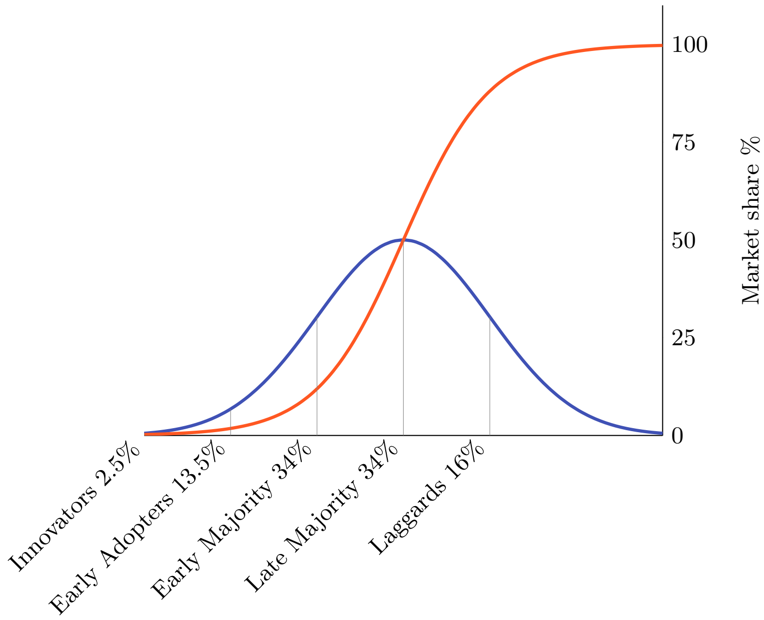


Figure 3.1: Diffusion of Innovation Curve by E. M. Rogers (1995). Time from left to right.

towards technology facilitating step 2: "persuasion that the innovation is positive or negative"). Thus, market segments can be separated into different target groups ranging from innovators that are very likely to know, try out, value or implement new technology to laggards with a high latency to adopt innovations (see figure 3.1). Especially step two, where people are persuaded to adopt a technology can be further analysed using the Technology Acceptance Model (TAM) of Davis (1989) as described in chapter 2. H. Lee, Hsieh and Hsu (2011) argued that IDT and TAM are not directly related but share one essential characteristic: both propose that the probability of accepting a technology or innovation is higher when people perceive it as useful and easy to use.

The construct of innovation also depends on its object and context. Significant confusions about proper definitions and relevant examples of what is genuinely innovative might stem from referring to different forms of classification. For example, Johannessen et al. (2001) argue that the degree of innovation depends on the context the innovation should be introduced into and distinguish between different objects for innovation: new products; new services; new methods of production; opening new markets; new sources of supply; and new ways of organising. In line with this, Norman and Verganti (2014) emphasise the importance of the context and, as parts of the context, propose socio-technical systems; ecosystems; business-models; products; services; processes; organisations; and organisational arrangements. They state that innovations might be classified by in its primary driver (technologies, markets, users, design).

For the field of HCI, new products and services are seen as the most critical domains

(Hassenzahl, 2004; Norman & Verganti, 2014). Also, the key characteristics of innovation as described above (newness and market penetration) must be thoroughly considered when designing deliberately for innovation. However, while objective indicators exist for market penetration and success of innovation (e.g. the number of users, market share, revenue), newness is more difficult to quantify. If innovative things are new – concerning novel and unique – and penetrate the market: when is something new enough to count as innovative and who are the important groups to judge this?

3.2 Perceived Newness and Perceived Innovation

E. M. Rogers (1995) equates innovation with perceived newness. So the answer to the question "when is something new?" is shifted to "when is something *perceived* as new?". Sethi, Smith and Park (2001) also focus on perceived newness, but also the perceived meaningfulness for innovation. Especially the perceived newness – typically refered to as the degree of unexpectedness, atypicality, obscurity, ambiguity, complexity, and uncertainty (Förster, Marguc & Gillebaart, 2010) – is at the core of innovation (Vogt, 2013).

Perceived newness is needed for innovation and has been shown to be an essential predictor of aesthetics in design (Crilly, Moultrie & Clarkson, 2004; Hekkert, Snelders & Wieringen, 2003). In line with this argumentation, Hollins and Pugh (1990) underline the importance of visual appearance and visual design for communicating innovation: "what-ever the product, the customers see it first before they buy it. The physical performance comes later, the visual always comes first" (p. 92). Radford and Bloch (2011) stress the point that visual design is crucial for the adoption of innovation. They argue that any functionally innovative product will eventually not be successful on the market as long as it does not communicate innovative characteristics in its visual design. For example, when a technologically groundbreaking new car does not convey innovative aspects in its visual design, it will not be perceived by customers as such. The difference between functional innovation and perceived innovation is fundamental: before a user can recognise competitive newness on the functional level, visual design is one of the few indicators for judging the degree of innovation of a product. To have a chance to be innovative, products and HCI must first be perceived as innovative.

In the context of HCI, Hassenzahl (2003) argues that *innovative*, *original*, or *new* are subjective, but highly desirable attributes for products and user interfaces. Scoring high on these dimensions can be achieved through functionality, context, interaction style – or presentation. This statement is in line with the claims of Radford and Bloch (2011) and Hollins and Pugh (1990) that the presentation of a product plays a central role in its adoption. Since innovation per se is based on newness, perceived innovation should also be based on newness.

Even though perceived innovation and newness are essential and desirable constructs for the field of HCI, a precise definition is missing. In this work, perceived innovation is understood as primarily focused on perceived newness (based on the definition of innovation by E. M. Rogers, 1995):

"Perceived innovation is the degree to which some individual attributes newness and novelty to a product or user interface."

This definition leaves out market success, one key element in definitions of innovation in the economics or organisational disciplines (Baregheh et al., 2009). Since future market success is difficult to predict and might be influenced heavily by inter-individual differences, this dimension is not included in the definition as used in this thesis. As proposed by Hassenzahl (2003), the perceived newness of a product or user interface might fulfill the need for stimulation of the user, which is beneficial for the product. The stimulation level of a product, service or user interface can be based on, for example, its "challenging and novel character" (Hassenzahl, 2004, p. 322). Measuring the perceived innovation of a user interface should follow these theoretical foundations.

Several approaches are available to measure perceived innovation. Qualitative content analysis has been argued to be not standardised and thus the comparability between different products is limited (Im & Workman, 2004). Quantitative questionnaires are often applied for measuring perceived innovation in the field of HCI (Hassenzahl, 2004; Hurtienne, Klöckner et al., 2015; Radford & Bloch, 2011). Adopting single Likert-scales for measuring perceived innovation is not optimal, since this approach might leave out essential facets and Measuring perceived innovation is more than analysing ratings of how "innovative" a product is perceived by participants (Im & Workman, 2004; Vogt, 2013). Instead, questionnaires that focus on perceived innovation cover a variety of adjacent constructs. For example, the subscale hedonic quality - stimulation of the AttrakDiff2 represents a semantic differential items such as "lame vs compelling", "dull vs creative", "harmless vs challenging", but naturally also "conservative vs innovative" (Hassenzahl, 2004). Other questionnaires on the perceived innovation are available. For example, the questionnaire of Im and Workman (2004) includes items like "This product is out of the ordinary" or "provides radical differences to industry norms" (p. 128). However, these often focus on industry processes or organisational innovation (Im & Workman, 2004; Vogt, 2013). Taken together, the construct of perceived innovation is highly subjective and the AttrakDiff2 provides a standardised and validated instrument for measuring perceived innovation, which is a) feasible and b) already focused on the HCI context.

3.3 Innovativeness as a Personal Trait

However, perceived newness and the disposition to adopt an innovation are likely to be influenced by inter-individual differences. For this, Goldsmith and Hofacker

(1991) suggest the term of innovativeness, the degree to which an individual tends to adopt an innovation, which differs profoundly between individuals. Following the above-described Diffusion of Innovations Theory (E. M. Rogers, 1995), individuals in the market can be classified into different segments that reflect their innovativeness: innovator, early adopter, early majority, late majority, laggard. Innovativeness has also been defined by Steenkamp, Hofstede and Wedel (1999) as a "predisposition to buy new and different products and brands rather than remain with previous choices and consumer patterns" (citet after Roehrich, 2004, p. 671). Roehrich (2004) extracted four sources from the literature that can explain this predisposition: 1) stimulation need, 2) novelty seeking[1], 3) independence towards others, and 4) need for uniqueness. Venkatesan (1973) showed early that people differ in their demand for stimulation and new products that are perceived as innovative can satisfy this demand. Importantly, this view distinguishes innovation (which includes market success) from perceived innovation: as long as the product is seen as innovative by an individual that seeks innovation, this need is satisfied regardless of the actual newness. Perceived innovation is a very subjective construct that describes the stimulation an individual experiences from a product (Hassenzahl, 2004). The need for stimulation is accompanied by seeking novelty. Adopting new products is a way for an individual to achieve both, and Hirschman (1980) even states that "novelty seeking is conceptually indistinguishable from the willingness to adopt new products" (p. 285).

Different instruments have been proposed to measure innovativeness. For example, Raju (1980) proposed a scale on innovativeness, which highly correlates with sensation seeking. Goldsmith and Hofacker (1991) developed a scale that focuses on domain-specific innovativeness (which is for them the "tendency to learn about and adopt innovations within a specific domain of interest"). Correlations of the scale with the following new-product-purchase range from .38 to .63 which shows its high predictive validity. In line with Hassenzahl (2004) and the concept of a product's hedonic qualities that stimulate the individual, Roehrich (1995) sees two main constructs at the core of innovativeness: a need for stimulation and need for uniqueness. His proposed scale, therefore, builds on two dimensions of innovativeness, namely hedonic innovation (e.g. "new products excite me") and social innovation (e.g. "I know more on new products than others"; translated in Roehrich, 2004).

In sum, the need for stimulation has a high impact on technology adoption and achieving perceived innovation and newness are essential for real innovation. Since perceived innovation is based on perceived newness, the question arises if individuals differ in their perception. Little research on age effects on perceived innovation and newness has been reported in the literature.

Though, for younger and older adults, such effects can be hypothesised. As described in chapter 2, older adults are less likely to adopt new technology compared to younger

[1] While 1) stimulation need focuses on the optimal level of stimulation for an individual (regardless of seeking to maintain it or not), 2) novelty seeking describes the active seeking of novelty. Therefore, stimulation need can, but does not necessarily lead to novelty seeking.

adults (Czaja et al., 2006; Olson et al., 2011) and prior experience with technology influences perceived newness (Crilly et al., 2004). Following the link between novelty seeking and innovation adoption, one explanation for this could lie in a reported decrease in novelty seeking during ageing (Zuckerman, Eysenck & Eysenck, 1978). However, Reio and Choi (2004) showed in a large sample size that older adults in fact only descend in sensational external types of novelty seeking (e.g., travelling to unfamiliar places or risky activities). Still, they remain stable or even increase in sensational internal novelty seeking (e.g. enjoying feelings and experiences) and cognitive novelty seeking in general (external: interacting with and understanding novel artefacts; internal: solving psychological problems or having creative ideas; Zuckerman, 1979, p. 140). Thus, the need for stimulation and novelty is – contrary to public opinion and stereotypes – not vanishing during ageing but instead shifts its focus. Innovation is essential also for the target group of older adults (Herstatt et al., 2011). Even though older adults are demanding clear benefits of novel technology to consider them for adopting (Melenhorst et al., 2006; Mitzner et al., 2010), stimulation by perceived innovation and newness is relevant also for the group of older adults.

In sum, perceived innovation and newness are desirable attributes of HCI, and it should communicate the degree of innovation it stands for on the functional level. Additionally, younger and older adults inhere a need for stimulation that can be satisfied via perceived newness of products and user interfaces. Visual design and interaction design are essential approaches to demonstrating the product's newness and innovation.

3.4 Interim Conclusion: The Design Problem

It is time to pull strings together and make the challenge explicit. Even though the exact wording differs over frameworks and definitions, innovation requires two things in its essence: a) novelty and perceived newness and b) perceived meaningfulness to facilitate market penetration. The focus of the present work will be on perceived newness. Still, perceived meaningfulness is not completely left out. Since the perceived meaningfulness of a technology also depends on its is ease of use (Davis, 1989), which will be addressed in the user studies. By this, the present work aims at HCI that is "most advanced yet acceptable" (MAYA, see also Hekkert et al., 2003).

To achieve perceived newness, user interfaces are likely to differ from existing and user interfaces making them less familiar to users (Crilly et al., 2004). Thus, innovative user interfaces can exclude older adults who a) might not have the cognitive abilities to learn innovative and unfamiliar user interfaces and b) often cannot profit from prior experiences with technology and transfer relevant prior knowledge to the new user interface. Age-inclusive design aims for user interfaces that allow users regardless of their age to use and interact with technology. To achieve this, an age-inclusive design should draw on prior knowledge that is shared between age groups. How can this be achieved?

The term “design” stems from the Latin “de” + “signare”, which means “making something, distinguishing it by a sign, giving it significance, designating its relation to other things, owners, users, or gods” (cited after Norman & Verganti, 2014, p. 80). Design tasks often involve so-called “wicked problems” that can also be defined as ill-structured (Rittel & Webber, 1973). Compared to “tame” problems that have a well-defined problem, a definite goal and a set of rules to work with (such as the tower of Hanoi puzzle, Coyne, 2005), wicked problems are loosely formulated, and no obvious methods can be applied. They are difficult to standardise, under-constrained, and a perfect “solution” for all its implicit by-problems is unlikely.

The design problem underlying this work is a combination of two wicked problems: user interfaces that are a) perceived as innovative while b) being age-inclusive. Approaches to age-inclusive interaction design were described in chapter 2. The next section summarises methods for innovative interaction design.

3.5 Designing for Innovation

Interaction designers can draw on a variety of methods, processes and techniques to stimulate innovative designs (Dalsgaard, 2014; Kelley et al., 2013). Stimulating creativity and designing for newness are tightly connected and the terms are often used almost interchangeably (Dark Horse Innovation, 2016; Kelley et al., 2013). Methods that drive innovation can be separated into at least two levels. On the macro-level, processes help project teams to structure their work and steer its outcome towards innovation. On the micro-level, specific methods can help individuals to stimulate creativity and brainstorm ideas. Both levels can be useful for the development of artefacts that users perceive as innovative. These two levels are described briefly in the following.

3.5.1 Macro-level: Innovation Processes

In the recent past, designing for innovation has been closely associated with design thinking, representing a toolbox of numerous methods and design strategies that have arisen in the last decades (Dalsgaard, 2014; Johansson-Sköldberg, Woodilla & Çetinkaya, 2013). Despite a long tradition in design disciplines and academic research, design thinking grew popular between 2005 and 2009 (Johansson-Sköldberg et al., 2013) and must be thoughtfully distinguished from “designerly thinking” (Goldschmidt & Rodgers, 2013). In its essence, design thinking is “a way of thinking that parallels other ways of thinking – like science thinking – but offers a way of approaching issues, problems and opportunities almost uniquely suited to innovation” (Owen, 2006, p. 3). Alternatively: “Design thinking is the way designers think: the mental processes they use to design objects, services or systems, as distinct from the end result of elegant and useful products” (Dunne & Martin, 2006, p. 517).

It is a collection of methods and techniques that facilitate particular ways of creative thinking while at the same time highlighting the importance of the target group a solution is developed for. Usually, design thinking draws iteratively on two forms of thinking: divergent and convergent thinking (Holtzblatt & Beyer, 2017; Kelley et al., 2013). In phases of divergent thinking, the design space is broadened and extended without constraining creativity to a set of given requirements. Only in phases of convergent thinking, constraints are used to narrow down the variety of possible solutions to practical ones that have the potential to solve the design problem effectively and efficiently: Design thinking is still a problem-solving activity in the end (Johansson-Sköldberg et al., 2013).

Even though design thinking includes both divergent and convergent thinking, its main contribution lies in divergent thinking that allows stimulating creativity in the development process. Thus, it provides a methodology to drive new solution to ill-defined problems and thus innovation and newness. Even though design thinking has been claimed to boost innovative solutions satisfying the user's need for stimulation (Hassenzahl, 2004), its explicit guidance for interaction design and visual presentation of user interfaces remains confined (Dalsgaard, 2014; Holtzblatt & Beyer, 2017).

Importantly, design thinking (as a general approach to innovation but not specific to HCI) is something else than HCI design and interaction design (Dalsgaard, 2014). As described above, innovation has become a necessity for many markets that are related to HCI and traditional requirements engineering has been found to be not suitable for driving innovation (Maiden, Gizikis & Robertson, 2004). In HCI design, Human-Centered Design (HCD; also called User-Centered Design, UCD) has been suggested as a successful innovation method that does not only drive creativity but also keeps solutions in line with the user's needs, the task's requirements and the context's constraints (Norman & Verganti, 2014). For example, Holtzblatt, Wendell and Wood (2005) provided a methodology for designing for a given context. Due to the introduction of smartphones and user experience gaining importance in the last years, this approach has later has been updated but remains stable in its essence (Holtzblatt & Beyer, 2017): including potential users and the context of use combined with iterative user testing in early phases of development.

Involving the user in the design process has become the industry standard for product development and user interface design. However, radical innovations also emerge without the involvement of users in early stages by using the approach of "genius design" (Saffer, 2010). User involvement has even been suspected to inhibit innovation (Y. Rogers, Sharp & Preece, 2011, p. 327): when the designer is exposed to and confronted with the user, it might inhibiting their creativity and potential for innovative ideas because the designer only focuses on the users' behaviour and tradition. However, involving users in an early stage can, in fact, promote innovation and creativity. It can be a deliberate tool to inspire designers with new forms of interaction design or visual design (Holtzblatt & Beyer, 2017; Norman & Verganti, 2014).

Even though the HCI community has well adopted this approach, the potential for innovation of this approach has been questioned, and functional innovation is in fact often driven by technological advances (Hasan & Yu, 2017; Norman & Verganti, 2014).

However, design thinking and Human-Centred Design can increase the perceived innovation while following only paths that lead to – from the perspective of potential users – useful and usable solutions. Since innovation, newness and creativity are often interwoven (Dalsgaard, 2014; Norman & Verganti, 2014), applying creativity methods in the design process can – but does not have to – enhance the chance of finding innovative solutions.

3.5.2 Micro-level: Creativity methods

To stimulate innovation and creativity specifically in interaction design and on the user interface level, numerous creativity methods have been suggested in the literature (Biskjaer et al., 2010; Dalsgaard, 2014; Hanington & Martin, 2012; Kelley et al., 2013; van de Sand, 2017). Instead of enlisting these – often not comparable – methods in detail, this section focuses on the useful framework of Biskjaer et al. (2010) that can help to structure methods and understand similarities and differences.

The authors clearly distinguish between creativity and innovation: "creativity broadly refers to the generation of novel approaches or ideas; innovation refers to the application of ideas in a specific context, often in the development of a specific product or service, and as such creativity is a pre-requisite for innovation" (Biskjaer et al., 2010, p. 1). Their framework facilitates the discussion and comparison of different creativity-oriented design methods according to four dimensions: 1) tradition and transcendence, 2) convergence and divergence, 3) degree of structure and 4) sources for inspiration. These dimensions can support the evaluation of a design method that aims at stimulating creativity and innovation.

Tradition and transcendence Design always had to balance tradition (i.e. the current state) and transcendence (i.e. a desirable future state) and even though the design is rooted in traditions (e.g. a given context, culture, prior experiences of people) it aims at "transcending them by anticipation and construction of alternative futures" (Ehn, 1988, p. 3). The first dimension, therefore, describes the tension between innovating the status quo and still driving change not too far away from it. This balancing act also applies to the problem of designing for innovation and age-inclusion.

Convergence and divergence Oscillations between convergent and divergent thinking have also early been regarded as an optimal general problem-solving strategy (Guilford, 1967). The second dimension is based on Löwgren and Stolterman (2004), who emphasise the clear advantages of alternating phases of broadening

Table 3.1: Overview of creativity methods in HCI (based on Biskjaer et al., 2010).

Method	Tradition / Transcendence	Convergence / Divergence	Degree of Structure	Sources of Inspiration
Brainstorming	rather tradition	divergence	loosely	none
Inspiration Cards	rather transcendence	both	loosely	many
Metaphorical Design	transcendence	both	moderately	few (but highly important)
Fictional Inquiry	transcendence	convergence	highly	few (one narrative frame)
Extreme Characters	transcendence	divergence	moderately	few (but highly important)
Future Workshops	both	divergence then convergence	highly	none
The Five Obstructions	traditional	convergence	highly	none to few

up and narrowing down the design space to stimulate creativity and innovative solutions.

Degree of structure In their third dimension, design methods aiming at innovation and creativity widely differ in their potential for both organisation of the design process as well as collecting tradition and competence within the team, factors that have been identified as two of the key benefits of design methods (Löwgren & Stolterman, 2004).

Source of inspiration Fourth and finally, a design method should be evaluated regarding its potential of providing stimulation and inspiration to the designer. Prior knowledge is as a fundamental part of creativity and knowledge about previous designs and ideas can inspire new ones (Eckert & Stacey, 2000). In line with this, Kelley (2001) report on "how designers at IDEO in a systematic way collect gadgets and materials to store them in a file cabinet for later use as sources of inspiration in subsequent design projects" (as cited in Biskjaer et al., 2010, p. 8).

The proposed framework can be applied to a variety of different design methods and is suitable for comparing them on the four dimensions suggested as necessary in the literature. Importantly, methods differ in their provided structure to the design team from loosely (Brainstorming) to strongly (Fictional Inquiry), their purpose of converging (Fictional Inquiry) or diverging (Extreme Characters), and the sources of inspiration from none (Brainstorming) to many (Inspiration Card Workshop). Additionally, the dimension of resources and effort needed for conducting the method is not included in the framework but must be carefully considered when comparing design methods in general and creativity methods specifically.

Table 3.1 summarises illustrative examples that show the bandwidth of design methods that can be reviewed with this framework. Biskjaer et al. (2010) review these design methods focusing on innovation and creativity. Metaphorical design, for example, as suggested by Madsen (1994), is classified as a well-suited method for transcendence (i.e., innovative solutions) while providing only a few sources of inspiration. In the following the single methods are briefly described.

Brainstorming is one of the most popular standard tools for stimulating creativity and problem solving. However, even though widely used, its effectiveness for ideation has been discussed (Rickards, 1999). Basically, the design team receives a (design) problem and starts generating ideas following four basic principles (combination and generation of ideas; no comments; short time with a maximum of design ideas; free associations and fantasy solutions are encouraged).

Inspiration Cards can be found in several variations (Biskjaer et al., 2010; Dark Horse Innovation, 2016; Kelley et al., 2013). Usually, several forms of cards are used and combined as sources of inspiration. For example, Technology Cards represent specific technologies or their application. Domain Cards represent specific information about the current domain. Concept Cards represent single concepts that might match the project context or not to stimulate divergent ideas and discussions. Basically, the design team receives these cards.

Metaphorical Design introduces core metaphors to the interaction design process that should stimulate novel forms of interactions and visualisations. Basically, the design team receives one or several metaphors (e.g. “a library is a meeting place”).

Fictional Inquiries focus on fictional situations, artefacts or narratives to break up the design team’s fixations on current situations and solutions. Basically, the design team receives a story that must be filled with details (e.g. explain the organisation and purpose of a public school to marsian tourists).

Extreme Characters aim at violating basic principles of usability to stimulate discussions about unexpected visions. Extreme characters focus the design team on challenging or even impossible target groups of their design problem (e.g. a blind person for a virtual reality game) breaking up conventional approaches to interaction design. Basically, the design team receives a set of exaggerated personas.

Future Workshops are highly structured venues for involving participants in the design process. Usually, they consist of three phases. First, in the critique phase, current problems are discussed. Second, in the fantasy phase, visions are generated on how these problems might be addressed. Third, in the implementation phase, the visions are adapted to the concrete context and the visions are drilled down to realistic next steps. Basically, the design team moderates a workshop with external people (e.g. citizens in a urban planning project).

The Five Obstructions originate from the disciplines of art and film directing. However, in interaction design, the creative power of constrains imposed on the design team has been used in form of so-called decisive constraints that can accelerate the design process and lead to very creative approaches to interaction design (Biskjaer & Halskov, 2014). Basically, the design team limits the design space via a set of self-imposed decisive constraints (e.g. only three colours in the user interface or must work on a low-resolution display).

Other creativity methods have been reported and are used in interaction design. However, some methods are rather applied in art and not always for classical interaction design (Biskjaer et al., 2010). Contextual Design (Holtzblatt & Beyer, 2017) might be interpreted as a creativity method as well. However, it involves a variety of different methods and the single methods can be classified differently within this framework. In fact, the single steps of Contextual Design are deliberatively selected to iteratively focus on transcendence and diversion and afterwards again on tradition and convergence. However, most methods involved during the extensive process of Contextual Design offer a basic structure to guide the team through the iterations. Finally, few (if even none) sources of inspiration are provided by the method. Instead, the sources of inspiration will emerge during the phase of user research and artefacts found in the current context, culture and people will stimulate the project team most.

The effectiveness of different creativity methods has been rarely subject to experimental investigation, including only low sample sizes (Biskjaer, Dalsgaard & Halskov, 2017; Chulvi, Mulet, Chakrabarti, López-Mesa & González-Cruz, 2012; Linsey et al., 2011; L. C. Schmidt, Vargas-Hernandez, Kremer & Linsey, 2010; Worinkeng, Summers & Joshi, 2013). Experimental data from research on design processes and creativity methods are often based on measuring novelty, variety, quantity and quality with standardised procedures (Dinar, Summers, Shah & Park, 2016). However, a complete objective quantification of achieved creativity due to a specific technique might always be influenced by subjective perceptions of the evaluator. Importantly, research in the field of design often relies on the subjective ratings of designers which is often assumed to be a more important measure than objective indices because it provides direct insights into the later acceptance of the novel method (Dinar et al., 2016; Zimmerman et al., 2007).

In sum, designing for innovation includes opening the design space to find ideas and novel approaches to existing problems. On a process level, design thinking and Human-Centred Design are suitable approaches that facilitate designing for innovation. However, their potential for guiding newness on the user interface level is limited. On the interaction level, different creativity methods are available for driving innovation and inspiring new solutions and details. One crucial drawback of research conducted in the field of interaction design is that novel methods and approaches are rarely compared to other approaches in an experimentally valid setting (Biskjaer et al., 2017; Zimmerman & Forlizzi, 2014).

3.6 Summary

Innovation is vital for the field of HCI. In its essence, the target group perceives innovations usually as a) new as well as b) meaningful to penetrate the market. Literature provides several processes, methods and techniques to stimulate the creativity of the project team and interaction designers leading to innovation and newness. To, at the same time, ensure that solutions (e.g. user interfaces or products) are meaningful to the target group and, therefore, can penetrate the market, the perceived effort required for using it should be minimised. Thus, the usability of the user interface cannot be neglected.

Since innovation and newness bears the danger (or even necessity) of rendering user interfaces as unfamiliar, innovations might enlarge the digital divide between younger and older adults. As we saw in chapter 2, besides User-Centred Design, several guidelines are available for interaction design that supports age-inclusive HCI. Still, innovation and age-inclusiveness are challenging to achieve in one user interface. However, given the growing user population of older adults and the digital divide: how can innovation be tailored to the characteristics of both younger and older adults? Despite applicable methods for either age-inclusiveness or innovation, these two goals are not easily brought together in one design methodology. The next chapter will, therefore, introduce an approach that is suitable to stimulate interaction designers to facilitate innovative solutions while at the same time considering the users' cognitive abilities and prior knowledge. Importantly, new methods for interaction design are often not compared to standard procedures. This thesis will address this shortcoming by not only introducing and applying a new method for innovative and age-inclusive interaction design, but as well evaluating it against other standard procedures. In this thesis, the standard procedure will be User-Centred Design (and subforms like Contextual Design, Holtzblatt & Beyer, 2017), as one of the most important methods in the field of HCI that allows for designing for both age-inclusiveness (see chaper 2) as well as innovation.

Chapter 4

Interaction Design with Image-Schematic Metaphors

The idea of following metaphors in both user interface and interaction design has a long history (Biskjaer et al., 2010; Blackwell, 2006; Madsen, 1994; Neale & Carroll, 1997), but also has been criticised as "not only unhelpful, but harmful" by Cooper (1995). Pitfalls of metaphorical interface design are seen that they – if not used appropriately – provide misleading visual cues, do not scale well, degrade over time, and are sometimes overused (Erickson, 1995). However, many of the used metaphors (e.g., the "desktop metaphor") are in fact analogies that mimick elements of real world in terms of the digital world.

This chapter introduces a sub-type of metaphorical design that utilises so-called image-schematic metaphors (Hurtienne, 2011; M. Johnson, 1987). In contrast to conventionally used metaphors in interaction design, image-schematic metaphors reveal how people understand and represent abstract concepts (and, therefore, allow to tap into technology-independent prior knowledge), metaphors in the user interface are usually not based on existing mental representations of users. Because of this, considering image-schematic metaphors for interaction design offers a variety of advantages over conventional user interface metaphors.

In the following, the literature on image-schematic metaphors and their application in interaction design is reviewed. Specifically, the focus will be on their usefulness for innovative and age-inclusive interaction design. The chapter closes with identifying gaps in literature and deriving specific research questions for this thesis to facilitate innovative and age-inclusive interaction design.

4.1 Metaphors: Theoretical Foundations

The term "metaphor" can refer to different things. This section will provide the background to distinguish image-schematic metaphors, which are central for the approach of this thesis, from linguistic, conceptual and primary metaphors.

4.1.1 Linguistic Metaphors

Linguistic metaphors are a rhetoric style element to enrich communication (e.g. "tree crown" for "the top of the tree") or a figure of speech (e.g. "The starting gun has been fired" for "The project started"). The core of metaphors always lies in taking a word from its original context ("crown": the king's or queen's head; "starting gun": sports) and transferring it to another context (the tree's head; project management). Thus, Keneth Burke defined metaphor as "a device for seeing things in terms of something else. It brings out the thisness of that and the thatness of this" (K. Burke, 1962, p. 503). However, besides the function in communication, metaphors are sometimes more than rhetoric figures: so-called conceptual metaphors. Certain linguistic metaphors can provide hints to underlying conceptual metaphors that constitute mental models of abstract concepts.

4.1.2 Conceptual and primary metaphors

In the fields of philosophy and cognitive linguistics, George Lakoff and Mark Johnson developed the theoretical basis for understanding the mechanisms of metaphors and how they mirror the way we think (Lakoff & Johnson, 1980). Metaphorical meaning is the understanding of one idea (the target domain) in terms of another (the source domain). In their book "Metaphors we live by" (Lakoff & Johnson, 1980), they coined the term of *conceptual metaphors*. This term refers to understanding and thinking about a concept (e.g. love) by using knowledge about another concept (e.g. journey) as surfaced in linguistic metaphors like "Our relationship has hit a dead-end street." Knowledge and assumptions about the source domain are called entailments and help to understand the target domain by using familiar concepts from the source domain (a journey has a start, end, crossings and dead-ends). The transferability from knowledge about the source domain to the target domain can differ: Similarly to the domain transfer distance discussed in chapter 2 (Diefenbach & Ullrich, 2015), the distance between the source domain and the target domain can be short or long, separating metaphors further into "Nahmetaphern" and "Fernmetaphern" (Plett, 1983, p. 83).

Language provides many examples of linguistic metaphors that allude to conceptual metaphors. For instance, expressions that fit the conceptual metaphor KNOWING IS SEEING are pervasive when talking metaphorically about the concept of KNOWING: "the scales fell from my eyes", "I did not see that coming", "now it is clear", or "he hides something". These expressions borrow from the source domain of SEEING to provide a better understanding of the target domain of KNOWING.

Lakoff and Johnson (1980) stress that such metaphoric expressions can reveal how we understand and think about abstract concepts. Often, plenty linguistic expressions are available for each single conceptual metaphor and these linguistic expressions are often present across different languages (Kövecses, 2006). A subtype of conceptual metaphors are so-called *primary metaphors*. In primary metaphors, the concrete source and more abstract target domain are correlated in experience. For example, the conceptual association between SEEING and KNOWING can be found in concrete experiences where not seeing equals not knowing (e.g. when playing cards). Because the experience that motivates primary metaphors is a form of very basic prior knowledge, they manifest in language across different cultures and age groups.

Thus, conceptual metaphors, and primary metaphors in particular, provide insight into the cognitive structure that is likely to be representative of a large group of people (Lakoff & Johnson, 1980, 1999). This relationship has two vivid implications: first, conceptual metaphors and primary metaphors allow to link perception, cognition, action, language and mental representation of abstract concepts in one framework; and second, they "make perception more automatic and ease the energy required to understand" (Schmitt, 1983, p. 366).

As the human's conceptual system is thouroughly structured by metaphors, it is assumed that some concepts can only be desribed metaphorically (Lakoff & Johnson, 1980). Referring to Paul Tillich's statement "Everything one says about God is a metaphor" (Tillich, 1957; cited after Knoblauch, 1999), Dan Saffer extended this to "Nearly everything one says about a computer is metaphoric" (Saffer, 2005, p. 12). In interface design, the assumed benefit of exploiting metaphors lies primarily in "allowing us to take our knowledge of familiar, concrete objects and experiences and use it to give structure to more abstract concepts" (Erickson, 1995, p. 66). Conceptual metaphors, and especially primary metaphors, might meet this demand.

Still, too broad and complex metaphors (e.g., COMPUTER IS A PERSON) are of little use for designers (Cooper, 1995; Neale & Carroll, 1997), because they are too open and do not offer concrete guidance that constrain the design space. This is a limitation of user interface metaphors as well (e.g., ONLINE ORDER PROCESS IS LIKE SHOPPING IN A PHYSICAL SUPERMARKET), because it remains unclear which parts of the source domain (supermarket) should be exploited in the user interface (online shopping). Identifying the most critical elements of a metaphorical mapping for interaction design is indeed challenging. Many metaphors can be boiled down and structured in a way that only the most important underlying implications and entailments of a metaphor become visible. For this task, so-called *image schemas* are a useful tool.

4.1.3 Image-schematic metaphors

M. Johnson (1987) introduced the term in reference to basic and re-occurring patterns that are extracted subconsciously while interacting with the physical world. For example, we perceive objects as being in- or outside of CONTAINERS (e.g., water in a bottle, coffee beans in a bag), as being UP in the air or DOWN on the ground, or as

Table 4.1: List of image schemas according to Hurtienne (2011)

Group	Image Schemas
BASIC	OBJECT, SUBSTANCE
SPACE	CENTER-PERIPHERY, CONTACT, FRONT-BACK, LEFT-RIGHT, LOCATION, NEAR-FAR, PATH, ROTATION, SCALE, UP-DOWN
CONTAINMENT	CONTAINER, CONTENT, FULL-EMPTY, IN-OUT, SURFACE
MULTIPLICITY	COLLECTION, COUNT-MASS, LINKAGE, MATCHING, MERGING, PART-WHOLE, SPLITTING
PROCESS	CYCLE, ITERATION, SUPERIMPOSITION
FORCE	ATTRACTION, BALANCE, BLOCKAGE, COMPULSION, COUNTERFORCE, DIVERSION, ENABLEMENT, MOMENTUM, RESISTANCE, RESTRAINT-REMOVAL, SELF-MOTION
ATTRIBUTE	BIG-SMALL, BRIGHT-DARK, FAST-SLOW, HARD-SOFT, HEAVY-LIGHT, SMOOTH-ROUGH, STRAIGHT, STRONG-WEAK, WARM-COLD

actively forcing other objects to move (COMPULSION). These basic mental concepts or image schemas (e.g., IN-OUT, UP-DOWN, COMPULSION) are – already in childhood – grounded in physical experiences (Mandler, 2010) and later also instantiated in our language. Image schemas are seen as the building blocks for many concepts in our world and help us to give structure to our environment. Hurtienne (2011) distinguishes 47 different image schemas as summarised in Table 4.1.

Often subconsciously, we use image schemas to describe both physical experiences as well as more abstract, non-physical concepts, through metaphors. The metaphorical mapping of abstract concepts to image schemas is called image-schematic metaphor (M. Johnson, 1987). Image-schematic metaphors are part of the group of conceptual metaphors but include only metaphors with an image schema as the source domain. Thus, the formulation of image-schematic metaphors always follows the pattern [ABSTRACT CONCEPT] is [IMAGE SCHEMA].

In some cases, the source and target domain in image-schematic metaphors correlate frequently in experience and form a primary metaphor. For example, QUANTITY and UP-DOWN are closely related in the physical world and have a rich experiential basis. More paper sheets on a pile (MORE) imply that the pile is higher (UP). More water in a glass of water (MORE) implies a higher water line (UP). The experiential basis for this primary metaphor is so frequently encountered that the link between QUANTITY and UP-DOWN is not only applied to the physical world but also extended to the quantification of more abstract concepts. For example, "a rising water line" and "a rising inflation" ground on the same metaphor: MORE IS UP – LESS IS DOWN. Interestingly, knowledge about such associations can be regarded as universal across cultures and age-groups because it stems from basic sensorimotor experience that can be frequently encountered during lifetime experiences. Also, metaphors that do not stem from sensorimotor experiences, but are acquired through language, can influence the way we think and act. Slepian and Ambady (2014) taught participants novel metaphors about the relationship between time and weight that were not based on prior physical experience.[1] Even though participants did not have any experiential basis for these metaphors, weight-ratings for books with the same weight but different publishing years on the envelopes strongly depended on the taught metaphor.

The claim that image-schematic metaphors as parts of mental models influence the way we think and act has received much empiricial support in various domains such as in cognitive linguistics, social psychology and HCI (e.g. Casasanto & Boroditsky, 2008; Hurtienne, 2017a; Löffler, Arlt, Toriizuka, Tscharn & Hurtienne, 2016; Macaranas, Antle & Riecke, 2012). For example, Löffler et al. (2016) could show that physical properties of tangible user interfaces influence the abstract meaning attributed to them. In their study, participants perceived heavier objects to be more important than lighter ones, which is in line with the image-schematic metaphor IMPORTANT IS HEAVY (e.g. "heavy matters of state").

The advantage of following image-schematic metaphors in interaction design compared to conceptual metaphors and even more broad user interface metaphors is that image-schematic metaphors constrain the source domain to image schemas. For example, in the desktop metaphor (MAIN SCREEN IS DESKTOP), the main point is not about mimicking an exact desktop in the user interface. Instead, it is about having a SURFACE where OBJECTS can be placed on top (UP). Also, conceptually putting OBJECTS into CONTAINERS is more important than illustrating an exact copy of a folder. The so-called invariance hypothesis states that not all but only some parts of the original meaning of the source domain of a metaphors are mapped to the target domain (Kövecses, 2006). The invariance hypothesis can, therefore, explain that image schemas and image-schematic metaphors sometimes allow decomposing complex metaphors into single ones that allow grasping the mental model more easily.

Based on similar assumptions, the framework of *blended interaction* also incorporates image schemas in interaction design (Jetter, Reiterer & Geyer, 2014). Based on embodied cognition and cognitive linguistics, it proposes *conceptual blends* that can explain how humans can draw on real-world concepts when they interact with digital technologies. The authors follow the work of Fauconnier and Turner (2008) (that ground their work on Lakoff and Johnson as well) and assume that "highly complex concepts in our conceptual system are generated from less complex or basic-level concepts by metaphorically integrating or *blending* them" (Jetter et al., 2014, p. 1141).

[1] For example, participants were primed by reading text passages like: "The past carries particular weight for who you are today" for PAST IS HEAVY and "The decisions of your past carry no weight. It is your decisions today that define who you are, and you must hold the present with great care." for PRESENT IS HEAVY (Slepian & Ambady, 2014, p. 311).

More specifically, complex concepts are mentally assembled by combining prior knowledge from different Input Spaces that are themselves based on a so-called *Generic Space* - which might include similar mental building blocks that are also described by image schemas. In HCI, the approach of conceptual blending and blended interaction shares crucial aspects with the theoretical foundations of image schematic metaphors. For example, both approaches ground on breaking down complex concepts into smaller parts that can be managed more easily in interaction design. Also, both approaches state that image-schemas could provide the basic and technology-independent prior knowledge that can facilitate intuitive interaction with technology because it is directly grounded on everyday experiences with our environment. However, image-schematic metaphors a) focus more explicitly on the link between image schemas and abstract concepts, and thus b) might guide interaction designers more easily. Even though both frameworks (blended interaction and image-schematic metaphors) are closely related, this work will focus on the framework of image-schematic metaphors and its terminology. Still, the framework of blended interaction provides an interesting additional perspective on the theoretical foundations of this work.

In sum, the reason why image-schematic metaphors have already been frequently investigated in HCI is that they bear the potential to support interaction design. Specifically, they can facilitate to align the visual and logical representation in the user interface to the user's mental representation of abstract concepts. Incorporating image-schematic metaphors like MORE IS UP - LESS IS DOWN in the user interface can lead to more intuitive interaction (Hurtienne, 2011). Section 4.3 will enlarge upon the application of image-schematic metaphors in the field of HCI and the interaction design process. The main promise is that user interfaces will benefit from conforming with image-schematic metaphors since they act as a form of prior knowledge the user can tap into easily during interaction. Several studies show that it is possible and beneficial to provide interaction designers with concrete image-schematic metaphors and that this method increases intuitive use in different application domains like tangible interaction, graphical user interfaces and gesture based interaction (Antle, Corness & Droumeva, 2009; Hurtienne, 2011; Hurtienne et al., 2010; Löffler et al., 2016). Basically, as long as a set of image-schematic metaphors is available for the design phase, interaction designers can use them as guidance in the concept phase. But before we turn to their application, the next section focuses on different sources that allow for the identification if image-schematic metaphors.

Analysing metaphorical language for image schemas and image-schematic metaphors reveals fundamental mental models. In other words, by analysing linguistic metaphors, image-schematic metaphors can be unveiled. For example, the words "in" and "push" can be both used non-metaphorically (e.g. "coffee *in* a cup"; "to *push* a car") or metaphorically , pointing to an image-schematic metaphor ("*in* 1987..." or "I felt *pushed* to that decision"). Additionally, other sources can serve as the basis for the extraction of image-schematic metaphors (e.g. observations, see next section 4.2).

4.2 Sources for Image-Schematic Metaphors in HCI

Especially in the context of HCI, language is the most frequent but not the only source for image-schematic metaphors (Hurtienne, 2017a). Four primary sources for images-schematic metaphors are relevant in HCI: documented linguistic data, user utterances, observations, and existing user interfaces (Hurtienne, 2017a). However, even though it is possible to analyse user interfaces for image schemas, this approach does not necessarily reflect the mental model of the real user (rather of the user interface designer) and could, therefore, lead to invalid image-schematic metaphors of users. Following Hurtienne (2011), image-schematic metaphors extracted from user interfaces often do not match those extracted from user utterances, raising questions about the validity of this source (Asikhia & Setchi, 2016). In the following, only the validity and reliability of the first three categories (documented linguistic data, user utterances, observations) will be discussed and compared.

4.2.1 Documented linguistic data

The literature contains various lists of documented metaphors that already provide metaphors that researchers and practitioners can directly access and utilise. In their book, Lakoff and Johnson (1980) provide and discuss a variety of conceptual metaphors and Lakoff, Espenson, Goldberg and Schwartz (1991) enlisted on over two hundred pages conceptual metaphors with linguistic examples. Furthermore, the ISCAT database summarises over 200 image-schematic metaphors allowing for their efficient application in HCI projects (Hurtienne, 2017a). The majority of documented metaphors stems from linguistic introspection (Lakoff & Johnson, 1980, 1999), as well as from the analysis of large linguistic corpora (Deignan, 2006; Stefanowitsch, 2006) and experiments (Hurtienne, 2011). The main advantage of this source lies in its efficient use because the metaphors are readily accessibly and can be applied without extra effort for elicitation.

Even linguistic corpora that have not been analysed for image-schematic metaphors yet can serve as a source for interaction designers. However, these corpora require more resources for the extraction process of image-schematic metaphors and Hurtienne (2017a) emphasises that the extraction process from this form of data is prone to errors resulting in invalid metaphors. Project teams are often not trained in the difficult extraction process of image-schematic metaphors which endangers the validity of the extracted metaphors (Dodge & Lakoff, 2005). To support the extraction process from large corpora, Gromann and Hedblom (2017) proposed to combine methods of automatic semantic labelling and clustering to extract image schemas from natural language automatically. They used the publicly available corpus of the European Parliament (EuroParl) which includes up to 60 million natural words for each of the 21 European languages. The authors constrained their image schema analysis tool to the five image schemas CONTAINMENT, PATH, SUPPORT (not reported by Hurtienne, 2011), SURFACE, and UP-DOWN. The results are promising in the way that the overall precision for their approach is 80% (with manual extraction by the authors as the golden standard). However, a large part of the high precision is explained due to the identification of easy/obvious key words like "in" and "within" for CONTAINMENT.

Moreover, documented linguistic data often presents written language from, for example, newspapers or the web and, therefore, differs from natural language. Additionally, currently available corpora focus mainly on global languages (mostly English, but also Spanish and Portuguese; https://corpus.byu.edu) and do not necessarily cover new technological topics (e.g. the concept of "cloud services").

In sum, documented metaphor lists and large corpora are an efficient basis for image-schematic metaphors. However, these lists are not available for all abstract concepts of interest. But when an existing database like the ISCAT-catalog enlists project-relevant image-schematic metaphors, interaction designers can directly incorporate them in their user interfaces.

4.2.2 User utterances and spoken natural language

Extracting image-schematic metaphors from natural language, for example, elicited during user interviews or other forms of user research, has been shown to be a feasible approach in practice (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013). User-Centred Design often already implies methods that stimulate user utterances which can be easily recorded (Holtzblatt & Beyer, 2017). However, to extract image-schematic metaphors requires to transcribe the recorded natural language, which is time-consuming itself.

User utterances are the most prevalent source of image-schematic metaphors because they can be elicited for a specific domain of interest. Since user utterances are the most prominent and convenient way of extracting image-schematic metaphors in HCI, the following section describes previous projects using this approach in HCI-related projects. Table 4.2 provides an overview of published works that report on the analysis of user utterances for image-schematic metaphors.

Importantly, due to different research questions, published methodologies differ in their extraction process of image schemas and image-schematic metaphors. While Raubal and Worboys (1999) extracted only image schemas from user utterances describing physical facts (e.g. "there is a blockage due to a physical door" – BLOCKAGE), Hurtienne, Klöckner et al. (2015) and Winkler et al. (2016) extracted image-schematic *metaphors*. Thus, even though earlier work addressed image schemas (e.g. Raubal et al., 1997), image-schematic metaphors have been introduced only afterwards to the HCI community (Hurtienne & Blessing, 2007).

As no common standards exist, the extraction process depends to a certain degree on the experience of the person that extracts the image-schematic metaphors. To illustrate the room for interpretation left by the unstandardised approach, consider

Table 4.2: Previous work on extractions of image schemas and image-schematic metaphors from user utterances in HCI-related projects.

Publication	Domain of application	Example
Löffler, Hess, Maier et al. (2013)	photo management and order processing system	PHOTOS AND RELATED META-INFORMATION ARE MATCHING
Löffler, Lindner and Hurtienne (2014)	Energy consumption in rural Africa	ENERGY IS ENABLEMENT
Hurtienne et al. (2010)	Heating controls	WARM IS UP – COLD IS DOWN
Hurtienne (2011), study 6	Bookkeeping system	INBOX IS a CONTAINER
Hurtienne, Klöckner et al. (2015)	Acoustic entertainment at home	BROADCASTING IS SELF-MOTION ON A PATH
Maglio and Matlock (1999)	World Wide Web	IMPORTANT PAGES ARE CENTER
Raubal, Egenhofer, Pfoser and Tryfona (1997); Raubal and Worboys (1999)	Navigating in airports	TICKET COUNTER AND SIGN ARE LINKAGE
Wilkie, Holland and Mulholland (2010)	Music interfaces	A CHORD IS A CONTAINER FOR NOTES
Winkler et al. (2016)	Social interactions between cars	ADDING A PERSON TO A CONVERSATION IS PUTTING SOMETHING INTO A CONTAINER

an example by Asikhia et al. (2015). The authors extracted the image schemas BRIGHT-DARK and BLOCKAGE from the user utterance "I could not figure out quickly the features because of the poor colour used for the label". While the BLOCKAGE image schema might be justified (extracted from the word "not" as well as from the real interruption during the interaction process), BRIGHT-DARK is not used in a metaphorical, but literal way, as it is the real, physical brightness that is referred to. Another problem of extracting image schemas from user utterances is put forward by Raubal and Worboys (1999). From the utterance "And again, the yellow signs are good", the authors extracted the image schemas BALANCE, SCALE, COLLECTION and OBJECT. Staying with the sentence itself, it can be stated that many "signs" are a COLLECTION of many OBJECTS, but the extraction of image schemas BALANCE and SCALE are problematic because they are only assumed from the context (observations) and cannot be found in the utterance itself. Thus, it is essential to distinguish not only between image schemas and image-schematic metaphors (the mapping of these basic elements to more abstract ideas), but also between direct evidence from linguistic material (e.g. image-schematic words) and assumptions from the context (which is problematic and prone to subjective interpretations).

In sum, the extraction process using natural language is not yet standardised (Hurtienne, 2011, 2017a). However, despite discrepancies between projects, attempts have been made to standardise the extraction process of image-schematic metaphors from natural language. For example, Löffler, Hess, Hurtienne et al. (2013) provide a toolbox that contains keywords that can be used during the extraction process (see section 4.4.1). They recommend that the extraction process usually should involve more than one person (Löffler, Hess, Hurtienne et al., 2013) that first extract image schemas from single metaphorical words and later on define corresponding image-schematic metaphors (e.g., when an employee talks about a contract management software, in the expression "I take note of the date *on* the contract", *on* leads to the image schema SURFACE and the image-schematic metaphor CONTRACT IS SURFACE). While some recommendations are available (e.g. multiple extractors), a clear standardisation of the process of extracting image-schematic metaphors is missing (Löffler, Hess, Maier et al., 2013). Thus, studies are not consistent in their methodology of applying image-schematic metaphors, rendering the learnability of the method challenging for novices.

4.2.3 Observation

Observation is a straightforward, but time-consuming way to explore the mapping physical dimensions to an abstract concept and thus extract a valid image-schematic metaphor. For example, Bakker, Antle and van den Hoven (2009) and Bakker, Antle and Van Den Hoven (2012) observed spontaneous body movements of children instructed to represent abstract concepts. From the body movements, they extracted image-schematic metaphors, for example, LOUD SOUNDS ARE BIG – SOFT SOUNDS ARE SMALL or LOUD SOUNDS ARE UP – SOFT SOUNDS ARE DOWN. However, even though observation is definitely as a well-suited approach to elicit image-schematic metaphors in specific domains (e.g. embodied interaction), this approach has been described as very effortful and often not feasible in practice because it requires the long observation of – when aiming at inclusive design – various user groups (Hurtienne, 2017a). Also, subconscious mappings need to be made explicit by participants, which might not always be possible.

4.2.4 Comparison

Different sources can serve as the basis for image-schematic metaphors. When image-schematic metaphors are already documented for the domain of interest, e.g. based on corpora or the ISCAT-database (Hurtienne, 2017a), they can provide a fast and efficient starting point. User utterances as a source for image-schematic metaphors provides a balance between efficiency (linguistic data on project-relevant domains is oftentimes already available in projects applying a User-Centered Design process) and effectiveness (valid and relevant image-schematic metaphors). From the three sources, user utterances might, therefore, represent the most promising approach to extract image-schematic metaphors. Finally, observations can lead to inspiring image-schematic metaphors, but the process takes more time and might not be feasible in most HCI projects since data for extraction is – if even possible – more difficult to collect.

Asikhia et al. (2015) and Asikhia and Setchi (2016) address the question of how these sources differ in their contribution to achieving a more reliable and valid set of image-schematic metaphors. Comparing user utterances during an experiment, observation and interviews after an experiment, they conclude that direct observation and user utterances are more useful than post-hoc interviews for the elicitation.

Analysing user utterances – as the main approach for finding image-schematic metaphors – also raises the critical topic of the inter-rater reliability between different people extracting image-schematic metaphors. Specifically for image-schematic metaphors, Hurtienne (2011) reports on the inter-rater reliability between different extractors of the same linguistic material (study 5). On the example of analysing single sentences for FORCE image schemas, the author reports on a moderate agreement (i.e., the inter-rater reliability; Cohen's kappa = .59) between extractors. The agreement with the image schemas that were the basis of the sentence was regarded as substantial (i.e., the validity; Cohen's kappa = .71). In a second study, Hurtienne (2011) could replicate this finding in a more realistic context (study 6, kappa = .68). Even though a sufficient inter-rater reliability has been shown for novices with little training, the effectiveness and efficiency of the extraction process still depends on the experience of the persons who extract the image-schematic metaphors (Löffler, Hess, Hurtienne et al., 2013). Additionally, the transcription of recorded natural language and the extraction process of image-schematic metaphors is very time-consuming. For example, Löffler, Hess, Hurtienne et al. (2013) recommend to plan half a day per contextual interview only for the extraction process (the resources needed for transcribing not included) plus one day for defining the final set of image-schematic metaphors.

4.3 Applications in Interaction Design

Various HCI-related projects exploited image schemas and image-schematic metaphors. This section focuses on the question of how well empirical data meet the promises of integrating image-schematic metaphors in the design process.

4.3.1 IS-M Extraction

Studies significantly differ in the amount and diversity of users recruited for the collection of user utterances for following extraction of image-schematic metaphors (see table 4.3). Most studies exploited user utterances from contextual interviews or think aloud protocols as a basis for the extraction process. Also, studies differ in their output of the extraction process. While most studies revealed concrete image-schematic metaphors (Hurtienne, Klöckner et al., 2015; Hurtienne & Langdon, 2010; Löffler, Hess, Hurtienne et al., 2013; Löffler et al., 2014; Wilkie et al., 2010; Winkler et al., 2016), the results are often limited to image schemas (Asikhia et al., 2015; Hurtienne, 2011; Maglio & Matlock, 1999; Raubal et al., 1997; Raubal & Worboys, 1999). When reported, only one or two extractors were responsible for the final set of im-

Table 4.3: User samples of used for collection of natural language in central HCI-related studies. IS: image schemas. IS-M: image-schematic metaphors.

Publication	Participants (differing in)	Data collection (duration)	Extractors
Asikhia et al. (2015)	42 (n.a.)	Think aloud protocols (15 min)	2 Focus on IS
Hurtienne and Langdon (2010)	10 (age: 26-84 years)	Contextual interviews (n.a.)	1 Focus on IS-M
Hurtienne (2011), study 6	3 (experience with the task)	Contextual interviews (120 min)	2 (kappa = .68) Focus on IS
Hurtienne, Klöckner et al. (2015)	8 (age: 57-86 years)	Contextual interviews (60 min)	2 Focus on IS-M
Löffler, Hess, Hurtienne et al. (2013)	company 1: 2, company 2: 3 (n.a.)	Contextual interviews (60-120 minutes)	2 Focus on IS-M
Löffler et al. (2014)	6 (English vs. Kiswaeli)	Interviews (30-40 minutes)	1 Focus on IS-M
Maglio and Matlock (1999)	24 (web-experience)	Post-study interview (n.a.)	2 Focus on IS
Raubal et al. (1997); Raubal and Worboys (1999)	8 (experience with the task)	Interviews supported by pictures (n.a.)	1 Focus on IS
Wilkie et al. (2010)	3 (experience as musician)	Dialogue of musicians on a song (30 min)	n.a. Focus on IS-M
Winkler et al. (2016)	10 (age: 22-61 years)	Contextual interviews (45 min)	3 Focus on IS-M

age schemas and image-schematic metaphors. Additionally, the documentation of the extraction process itself is often limited in publications rendering the interpretation of the extracted image schemas and image-schematic metaphors challenging and making the process of extracting image-schematic metaphors less accessible for other researchers and practitioners.

The number of participants significantly depends on the methodology of the overall development process. Based on the user research phase of the User Centered Design process (e.g Holtzblatt & Beyer, 2017), a sample size of 4 to 12 participants[2] has been suggested to be sufficient as a basis for extracting image schema and image-schematic metaphors (Löffler, Hess, Hurtienne et al., 2013). However, it is harder to determine the dominant image-schematic metaphors when the sample is very small and too narrow.

4.3.2 Interaction design and evaluation

The method of image-schematic metaphors has received attention in the fields of tangible user interfaces, gestural interaction and traditional graphical user interfaces.

[2]Usually, in User-Centred Design, participants solve the same tasks similarly, thus more participants will reveal only little more insights.

Table 4.4: Comparison of studies using image schemas or image-schematic metaphors for interaction design or evaluation. IS-M: image-schematic metaphors

Publication	Prototypes	Evaluation	Main finding regarding IS-M
Antle et al. (2009)	1 prototype for embodied interaction (based on the BALANCE image schema)	Qualitative user study (45 participants)	Prototype perceived as novel; behaviour conforming with IS-M but participants cannot verbalise their mental model
Asikhia and Setchi (2016); Asikhia et al. (2015)	3 interactive products differing in extent of complying with IS-M	Quantitative user study (42 participants)	Match between IS-M of the product and IS-M extracted from user utterances increases intuitive use
Bakker et al. (2009), Bakker et al. (2012)	12 prototypes for embodied interaction (based on IS-M on music representation)	Quantitative user study (39 participants)	Spontaneous interaction with prototypes consistent with IS-M
Hurtienne (2011), studies 1-4	laboratory study, no complex prototypes but simple stimuli	Quantitative laboratory experiment (128 participants in total)	IS-M conform GUI-elements more intuitive to use
Hurtienne, Klöckner et al. (2015)	1 interactive prototype complying with IS-M	Quantitative user study (78 participants)	IS-M prototype descriptively intuitive to use, age-inclusive and innovative
Löffler et al. (2014)	8 single paper prototypes, consolidated to two main prototypes (IS-M vs. not)	Informal evaluation (participants: n.a)	n.a.
Lund (2003)	1 prototype complying with IS-M compared against baseline	Qualitative user study (16 participants)	Participants used the same IS-M in their language, regardless of concrete instantiations of IS-M in the prototypes
Wilkie et al. (2010)	2 interactive products differing in extent of complying with IS-M	Qualitative, two experts in musical user interfaces	Differences in compliance with IS-M between products
Winkler et al. (2016)	10 prototypes; half of them complying with IS-M, the other half industry standard	quantitative user study (20 participants)	IS-M prototypes more intuitive or comparable to industry standard

Table 4.4 summarises studies that used image schemas or image-schematic metaphors to guide or evaluate interaction design.

Hurtienne, Israel and Weber (2008) and Hurtienne (2011) laid essential methodological foundations and explored prerequisites for the application of image-schematic metaphors in graphical and tangible user interfaces. In a series of laboratory and field studies, Hurtienne (2011) addressed the questions whether image schemas and image-schematic metaphors can provide the basis for a practical and feasible interaction design method with the goal of designing for intuitive use. The results were promising and stimulated further research.

To further develop the methodology and explore potential for optimisations, Löffler, Hess, Maier et al. (2013) applied the approach of image-schematic metaphors in an industry project. Two teams worked on projects on software development. A basic User Centered Design process and user participation were introduced to both teams, but only one team additionally received image-schematic metaphors. As expected, the team with image-schematic metaphors stated qualitatively that the development process was more structured due to the image-schematic metaphors. However, both teams used User-Centred Design for the first time, which could have affected the assessment of the method of image-schematic metaphors compared to the baseline team. In sum, this study could not show how project teams would adopt image-schematic metaphors in their process because both teams learned a new method.

After integrating image-schematic metaphors into the User-Centered Design Process, Hurtienne, Klöckner et al. (2015) provide a set of recommendations for future applications of this method. Interestingly, they already focus on the tension between intuitive use, innovation and age-inclusiveness. In a user study with 72 participants, their developed user interface in the application domain of acoustic entertainment at home scored high on all of these three dimensions. However, these findings must be considered carefully. First, no baseline was designed and evaluated rendering the interpretation of the "above average"-scores difficult. Second, it was concluded that age-inclusiveness is given only on account of no statistically significant differences between age-groups, but no measurements for the age-related decline were applied. Third, only questionnaire data was used, and no behavioural measurements were applied. Still, in sum, this study provides first insights on an entirely new designed user interface based on image-schematic metaphors compared to a single re-design of an existing user interface (e.g. Hurtienne, 2011; Hurtienne et al., 2008).

Addressing the need for a direct comparison of prototypes that were developed explicitly with image-schematic metaphors versus not (industry standard), Winkler et al. (2016) report on a User-Centered Design process (including contextual inquiries with younger and older adults) in the application domain of social interaction between cars. Even though their final evaluation of the developed prototypes involved only 20 participants, prototypes that were based on image-schematic metaphors were perceived as more innovative – measured by the subscale "hedonic quality - stimulation" of (Hassenzahl, 2004) – than the industry standard while measures of intuitive use (QUESI, after Naumann & Hurtienne, 2010) and task performance (task completion rates and time) were at least equal or better for the image-schematic metaphor prototypes. At the same time, no differences between age groups were found in any variable.

Finally, Löffler et al. (2014) also integrated image-schematic metaphors in the design process. Despite a small number of participants in the design phase and an only qualitative evaluation of the developed prototypes, results support the claim that this method can facilitate the creation of intuitive interactions.

Image schemas and image-schematic metaphors have also been used as an inspiration for other interaction techniques. For gestures, Hurtienne et al. (2010) found a high agreement between participants when performing arm-gestures for abstract concepts as well as a high agreement between observed gestures and gestures that image-schematic metaphors would have predicted. Chattopadhyay and Bolchini

(2015) applied image schema (UP-DOWN and LEFT-RIGHT, but not image-schematic metaphors) as a framework for intuitive free-hand gestures and found that these can be useful for the definition and prediction of intuitive gestures. Antle et al. (2009) also applied image schemas as a framework for embodied and full-body interaction.

Contrary to the application of image-schematic metaphors as a guiding and prescriptive interaction design method, Asikhia et al. (2015) used image schemas to evaluate interaction with products regarding intuitive use. They propose a framework that compares image schemas, but not image-schematic metaphors, found during real interaction (extracted from user utterances, observation, interviews, questionnaires) with image schemas extracted from manuals (the designers' mental models). Their final score for intuitive use – *Asikhia's Q* – can be used to assess the match of intended and achieved image schemas. When the match is high (i.e., between designer's intended mental model and the user's real mental model), the intuitive use is also high. The major problem with this approach is that the validity of extraction image schemas from manuals cannot be taken for granted as a good predictor for the intended image schemas by the interaction designer.

Comparing these studies reveals three general approaches to evaluation. In the first approach, existing interactive products are evaluated based on their compliance with image-schematic metaphors (Asikhia et al., 2015; Wilkie et al., 2010). Even though this approach increases the probability that the investigated interactions are more representative and realistic for current interaction design, it remains unclear whether the products also differ in other aspects than confirming with a given image-schematic metaphor. Thus, a controlled experimental manipulation is difficult in this context and obtained results are difficult to interpret. Also, prescriptive recommendations for interaction design are difficult to derive from this approach.

The second approach focuses on single interaction elements (Hurtienne, 2011, study 1-4). By this, a standardised experimental manipulation is possible, and confounding variables are kept constant. However, the external validity and generalisability of these findings to more complex interactions is an open question.

Finally, the third approach investigates more complex interactions and are explicitly designed to comply with relevant image-schematic metaphors or not (Hurtienne, Klöckner et al., 2015; Löffler et al., 2014; Winkler et al., 2016). Developing interactive prototypes (paper-based or digital) combines the strengths of the first two approaches: experimental manipulation (while still not wholly standardised) and external validity (complex and interactive prototypes instead of single GUI-elements). However, studies in the third category suffer also from various methodological shortcomings. For example, Hurtienne, Klöckner et al. (2015) did not compare their prototype – that complied with the prescribed image-schematic metaphors – to a baseline, which makes the interpretation of their results difficult. Löffler et al. (2014) did not conduct a formal evaluation but relied on positive qualitative results with only a few participants. Winkler et al. (2016) compared prototypes complying with image-schematic metaphors or not for five use-cases, but the evaluation study included only few participants that were additionally screened only to a limited degree regarding cognitive abilities and prior technological knowledge.

Additionally, comparing interfaces that are based vs not based on image-schematic metaphors affects differences between interfaces on various levels (e.g., more visual than text-based presentation, the logical order of steps, functionality). Thus, in studies comparing only two prototypes, the danger of comparing something else than "complies with image-schematic metaphor or not" is high.

4.4 Synthesis: IS-M Methodology for HCI

Previous work attests that image-schematic metaphors can provide a useful framework for interaction design. A variety of studies contribute qualitative insights and guidelines for grasping this method, but it remains challenging to apply for researchers not familiar with it and not connected to one of the few teams researching this topic. The purpose of this section is therefore to review the current recommendations and to distil best practices from several projects into a state-of-the-art methodology. The here presented methodology will serve as a basis for the formulation of open research questions as well as for one of the contributions of this thesis: a better-documented methodology for utilising image-schematic metaphors in interaction design to design for innovation and age-inclusiveness.

Hurtienne, Klöckner et al. (2015) already provide a core methodology for practising image-schematic metaphors in the design process. More specifically, the authors named the following three stages, where image-schematic metaphors can be blended into the User-Centered Design process.

1. **User Research** User utterances collected during user research serve as the basis for extracting image schemas and image-schematic metaphors.
2. **Ideation** Image-schematic metaphors can be used as a design guidance for intuitive use.
3. **Prototyping & Testing** Image-schematic metaphors represent basic guidelines that support the design team in designing for the user's mental model on the interface level.

The effort required to integrate image-schematic metaphors in these three phases differs greatly (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013). Most resources are needed in the user research phase for extracting image-schematic metaphors from user utterances. The basic training the team in the new method occurs usually directly on the project, with an estimated time required for explicit training of two hours (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013). However, carrying out the concrete analysis of image schemas and image-schematic metaphors and finally providing the set of relevant and most dominant image-schematic metaphors has been shown to be very resource-consuming. Especially when the team consists of non-experts in this area, Hurtienne, Klöckner et al. (2015) recommend the addition of an expert of image schemas and image-schematic metaphors. However, as soon as the list of relevant image-schematic metaphors has been agreed on, the implementation in the design process is less time-consuming.

4.4.1 User Research: extraction of image-schematic metaphors

As summarised in section 4.2, user utterances are regarded as the most feasible and efficient source for extracting image-schematic metaphors. However, the extraction process is not yet standardised and difficult to carry out by novices. In sum, the literature suggests three steps for the extraction of image-schematic metaphors: 1) collecting user utterances, 2) extraction of image schemas and 3) extraction of image-schematic metaphors.

1. **Collecting user utterances.** The natural language of potential users constitutes the basis for the succeeding extraction process. Depending on the project context, the team can choose from conventional collection methods such as user interviews on or contextual inquiries (Holtzblatt & Beyer, 2017; Hurtienne, Klöckner et al., 2015). The user utterances need to be transcribed on a word-by-word basis, which is time-consuming (but can be outsourced).

2. **Extraction of image schemas.** Based on the collected user utterances, single image schemas are extracted. Since this is a non-trivial task for novices in this field, supervision is strongly recommended. Recommendations on the depth of extraction differ. To facilitate the extraction of image-schematic metaphors, only metaphorically and non-literal meanings should be extracted and ideally only for those parts of the transcripts that contain expressions on abstract concepts that are relevant for the project and design process (Hurtienne, Klöckner et al., 2015). Examples and keywords for this process can be found in Löffler, Hess, Hurtienne et al. (2013) (see figure 4.1). Maglio and Matlock (1999) also provide a list of keywords that can support the extraction of image schemas.

3. **Extraction of image-schematic metaphors.** Finally, the extracted image schemas are transformed into image-schematic metaphors by combining each found image schema with its abstract concept. Each image schema (e.g. COMPULSION) provides most insight when it is associated with a concrete functionality or abstract concept (e.g. TRANSFERRING MONEY BETWEEN TWO BANK ACCOUNTS). By this, the mental representation of an abstract concept is made explicit (e.g. TRANSFERRING MONEY BETWEEN TWO BANK ACCOUNTS IS COMPULSION). This can lead to a considerable number of image-schematic metaphors of which only a fraction is potentially applicable and relevant for the following phases of ideation and interaction design. Thus, image-schematic

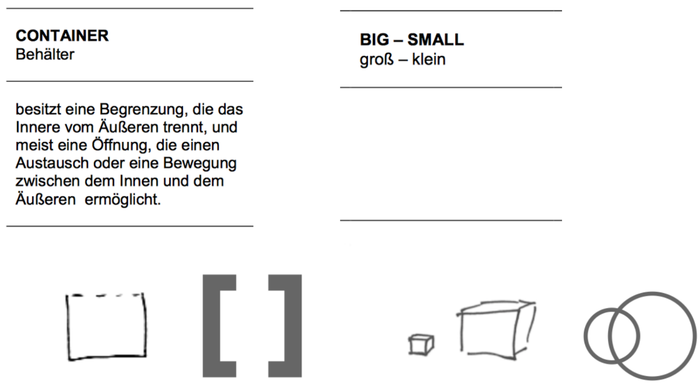


Figure 4.1: Examples for the definition of image schemas. Taken from Löffler, Hess, Hurtienne et al. (2013).

metaphors should focus on those abstract concepts that are of interest for interaction design of project-relevant functionality.

Several methodological issues in the phase of extracting image schemas and image-schematic metaphors to linguistic material have been subject to discussions and research. Contrary to extracting only metaphorical instantiations, Raubal and Worboys (1999) also extracted literal expressions containing keywords for specific image schemas (e.g., "I go forward": FRONT). The theory of image schemas is not constrained to metaphorical meanings in language. Still, image schemas alone without a connection to abstract concepts via image-schematic metaphors are of little value for interaction design that focuses on the mental representations of the users. Even though Raubal and Worboys (1999) adopt the theory of image schemas in a very applied project, their aim was rather to provide a formal model for the process of human wayfinding instead of investigating mental representations. Similarly, Chattopadhyay and Bolchini (2015) referred to the term image schemas while focusing only on the dimensions *left-right* and *up-down*. Even though these two dimensions are of image-schematic nature, this approach does not draw full potential of image-schematic metaphors uncovering the user's mental model. Image schemas alone are a promising start to describe HCI and, importantly, image schemas are very useful for describing, understanding and designing physical applications. Also, the essence of image schemas (e.g. FORCE image schemas) can best be understood when they are taught phyisically (Hurtienne, Löffler, Gadegast & Hußlein, 2015). But since the focus of designing with image-schematic metaphors primarily lies in achieving a good match between the user's mental representation and the interaction or interface, only metaphorical instantiations of image schemas should be considered for extraction (Hurtienne, 2011; Hurtienne, Klöckner et al., 2015).

The question of whether an expression is metaphorical has been addressed extensively. Following the definition of a metaphor by Lakoff and Johnson (1980), three prerequisites must be met for a word or phrase to be classified as a metaphor (Schmitt, 1983):

1. Non-literal meaning. The word or phrase can be understood non-literarily.
2. Source area. The word or phrase is rooted in an area of sensory or cultural experience.
3. Target area. The word or phrase is used in a second, abstract area.

To standardise the extraction of metaphors to a certain degree, Schmitt (1983) proposed an iterative procedure instead of solely relying on subjective extractions often based on gut feelings. The author recommends to define metaphors a priori or after screening the text and only afterwards to analyse the linguistic material in depth. This approach also relies on experience but helps the extractor to focus on promising and relevant phrases and sentences. Others have also proposed iterative approaches to the analysis of metaphorical phrases in the linguistic material which might increase the reliability of the process (Gibson & Zellmer-Bruhn, 2001). Additionally, when analysing natural language stemming from transcripts, Schmitt (1983) also recommends researchers responsible for extraction to participate in a self-interview and subsequently examine their language. By this, the researcher is made more cognizant of her own metaphorical expressions decreasing the probability of utilising them in the following interviews, which might bias the interview partner. This approach also shapes the skills needed for the later metaphorical classification of phrases and sentences. Schmitt (2003) also presumes that it might be easier to perceive metaphors that are striking, disruptive or do not correspond to one's sentiment. However, no empirical evaluation of this approach was reported.

To standardise the extraction process even further, Steen et al. (2010) propose a very detailed procedure to identify metaphors in discourse: MIPVU (abbreviation for Metaphor Identification Procedure of Vrije University). The basic procedure includes analysing the transcripts for possibly metaphor-related words on a strictly linguistic basis (Steen et al., 2010, p. 25). Guidelines support the decision of whether a word can be regarded as metaphor-related distinguishing between three different types (indirect, direct, implicit). Also, the analysis screens for signal words for cross-domain mappings (metaphor flags). Contrary to other recommendations (Hurtienne, Klöckner et al., 2015), only entire lexical units and no pre- or suffixes are analysed (e.g. under-pass). The approach – rooted in linguistic research – can structure the process to a high degree and standardises decisions on whether an expression is metaphoric or not. However, the extreme standardisation entails a very time-consuming process and might not be feasible in practice. Also, the focus of this method is not on identifying image-schematic metaphors and might be, therefore, necessary in linguistic studies but of limited use in the field of HCI.

Finally, Löffler, Hess, Maier et al. (2013) developed a more practical toolbox that supports non-experts in the extraction process (e.g., a list of German keywords indicating possible candidates for extracting image schemas and image-schematic metaphors).

Furthermore, they offer valuable insights into the current standard of the extraction process of image schemas and image-schematic metaphors. The process of extracting image schemas and image-schematic metaphors can be instructed and learned in a short time. Reported training sessions range from 30 minutes to half-day workshops (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013). Hurtienne, Löffler et al. (2015) report on two different approaches for introducing FORCE image schemas to novices. They found that tangible experiences were better suited to convey a basic understanding. Still, the validity and the reliability of the resulting metaphors have been rarely investigated. Löffler, Hess, Maier et al. (2013) and Hurtienne, Klöckner et al. (2015) therefore recommend to integrate an "image schema expert" in the development team. Especially the formulation of image-schematic metaphors has been reported to be a difficult task (Hurtienne, Klöckner et al., 2015; Winkler et al., 2016), a task which can and should be supported by an experienced expert.

Unfortunately, other publications do not provide much detail about their extraction methodology of image schemas and image-schematic metaphors (Asikhia & Setchi, 2016; Asikhia et al., 2015; Dodge & Lakoff, 2005; Raubal et al., 1997). Thus, the publications of Hurtienne, Klöckner et al. (2015) and Löffler, Hess, Maier et al. (2013) can be regarded as the current standard for the extraction process in HCI-related research and projects. Following their methodology promises to extract image-schematic metaphors that can act as a window to the user's mental models and representations of abstract concepts. Still, the effort required for a final set of domain-specific image-schematic metaphors is high. Specifically, two reasons decrease efficiency which must be addressed.

1. A large amount of linguistc data must be analysed for image schemas and image-schematic metaphors. For extractors, this is a repetitive and unfamiliar task and might render them less sensitive for more complex metaphors. It has been observed that extractors stick to most obvious keywords for image schemas (e.g. *in* for IN-OUT) but ignore more complicated image schemas like MOMENTUM that require to go beyond single signal words.
2. Concrete and reliable rules for the extraction process are missing and often not possible. Thus, extractors will differ in their implicit rules that lead to slightly differing extraction outcomes.

Possible solutions addressing the problem of efficiency can, therefore, begin in reducing the amount of parts of the transcripts that have to be analysed, for example, by removing all parts of the recorded linguistic data that are irrelevant for the project from the transcription and extraction process (Löffler, Hess, Hurtienne et al., 2013). Clear rules that allow at least partial automation would further support the extraction process, and databases that provide a pool of image-schematic metaphors would even render the time consuming extraction process unnecessary in some projects. However, automation and reliance on image-schematic metaphors reported in databases might imply that those metaphors that are most specific to the current project are not uncovered.

4.4.2 Ideation and prototyping with image-schematic metaphors

The previous section covered the extraction of image-schematic metaphors in the user research phase. However, the extraction does not contribute any benefit as long as it is not fed into the design process. In the next section, the major advantages of image-schematic metaphors for the ideation phase and the resulting user interfaces are discussed.

Primarily, grounding interaction design in image-schematic metaphors promises to enable the designer to align the system representation to the user's basic mental model of the domain while at the same time leaving room for innovative user interfaces. By this, younger and older adults can efficiently draw on the same prior knowledge which is less dependent on technological experience leveraging age-inclusive interaction. Importantly, the addition of image-schematic metaphors itself has been suggested to be an inspiring source for creative new designs (Hurtienne, Klöckner et al., 2015).

A valid set of image-schematic metaphors extracted as described in section 4.4.1 can facilitate the ideation and prototyping process twofold:

1. As a source of inspiration, image-schematic metaphors might stimulate creativity and lead to novel and non-conventional solutions (for creativity methods not related to image-schematic metaphors, see also Biskjaer et al., 2010).

2. Image schemas as a meta-language for interaction design can structure the communication within a team while stepping through the design process (Hurtienne, 2011; Hurtienne et al., 2008). This may facilitate and possibly accelerate the design process.

The literature reports on several approaches to integrate image-schematic metaphors in the phases of ideation and prototyping. (Hurtienne, 2011, study 7) and Hurtienne et al. (2008) proposed utilising the creativity method of a morphological box. This approach requires designers to come up with a set of diverse ideas or specific interface-solutions for each image-schematic metaphor from which later on only one is chosen and implemented. This is in line with the claim that creativity in interaction design mainly stems from divergent thinking and the exploration of a variety of different design solutions (Biskjaer et al., 2010). Hurtienne, Klöckner et al. (2015) explicitly refrained from applying this technique to allow designers a less constrained and more efficient design process.

Unfortunately, other publications do not describe in detail how image-schematic metaphors were introduced to the interaction designers and integrated into the overall design process. Löffler, Hess, Maier et al. (2013) successfully blended image-schematic metaphors in a company's current development process, but they provide no further information about particular procedures (even though the authors presented a best practices guide in form of the IBIS-manual at the end of the project). Based on image-schematic metaphors, Bakker et al. (2009) developed tangible and embodied prototypes. Their procedure was twofold. First, they developed low-fidelity prototypes that matched previously identified image-schematic metaphors (here called embodied metaphors). Only in a second step, high-fidelity prototypes were developed. They also do not provide enough details to replicate the design process. In the same context as Bakker et al. (2009), Antle et al. (2009) address more in-depth the question of how image-schematic metaphors (which they also call embodied metaphors) are best to be used as a basis for interaction design. By comparing three different projects (Antle et al., 2009; Bakker et al., 2009, 2012), they distil similarities and reoccurring insights from various application domains and design processes to support researchers in their design process with image-schematic metaphors. Even though the methodology of all three projects follows the approach of Research Through Design (Zimmerman et al., 2007), their insights still contribute a substantial source of recommendations and guidelines that can be helpful in general for designing with image-schematic metaphors. Amongst other general insights across the three projects, they highlight two challenges for the process of translating metaphors into design ideas:

What? Each image-schematic metaphor can be applied to several aspects of the user interface. For example, a metaphor might be mapped to an action or input, output or even only the basic visual layout of the interface. The final decision of which part of HCI should be supported by the image-schematic metaphor is left to the interaction designer.

How? For each aspect of interaction, multiple design solutions might exist that are all in line with the image-schematic metaphor. Even though the metaphor guides the interaction designer, this often even enlarges the plausible design space instead of narrowing it down.

However, both challenges in fact represented as a positive observation, since image-schematic metaphors stimulate creativity and inspiration. Antle et al. (2009) thus provide many recommendations that extend the knowledge about designing with metaphors. First, they draw the conclusion that most conceptual systems are understood regarding one or more spatial primary schemata (e.g., up-down, in-out). Second, interactions that are based on image-schematic metaphors should not neglect the demand of discoverability. Immediate feedback, affordances and task-related cues are essential and well-established ingredients for a usable interaction. However, the problem of discoverability can also be regarded as inherent to full-body and gestural interfaces and is not directly following from incorporating image-schematic metaphors to the design of, for example, graphical user interfaces. Third, image-schematic metaphors should be instantiated in as many parts of the interaction as possible (input, output, layout). Finally, the authors urge to distinguish between different forms and goals of HCI artefacts. For example, the design process should focus on only single and not interfering image-schematic metaphors when the goal is good usability and ease of use, while for art installations, multiple image-schematic metaphors can be used in an unconstrained interaction space.

In sum, image-schematic metaphors should be a suitable method for stimulating interaction designers to come up with creative design solutions that are in line with the mental representations of potential users, resulting in efficient HCI. Image-schematic metaphors also might stimulate the interaction designer to come up with novel and innovative user interfaces. The next section will further sharpen the promises of image-schematic metaphors for age-inclusive and innovative interaction design.

4.5 IS-M for Innovation and Age-Inclusiveness

Implicitly, image-schematic metaphors inhere an outstanding potential for designing both innovative and age-inclusive user interfaces. The purpose of this section is, therefore, to connect the method of image-schematic metaphors to chapters 2 and 3, to make advantages for innovative and age-inclusive interaction design clear and to lay the foundation for deriving focused research questions in section 4.6.

4.5.1 IS-M and Innovation

Blending image-schematic metaphors into the design process can, in fact, lead to innovative interfaces (Hurtienne, Klöckner et al., 2015; Winkler et al., 2016). Innovative interfaces refer here not to innovation in its narrow sense – thus, as also requiring market success (Dahlin & Behrens, 2005) – but to perceived innovation (Hassenzahl, 2004; Johannessen et al., 2001; Radford & Bloch, 2011). What are the mechanisms that enable image-schematic metaphors to stimulate creativity and finally increase perceived innovation? In the following, the framework of Biskjaer et al. (2010) will serve as a tool for illustrating the potential of image-schematic metaphors for innovative user interfaces on the dimensions of *tradition and transcendence*, *convergence and divergence*, *degree of structure*, and *sources for inspiration* (see chapter 3).

Interaction design based on image-schematic metaphors oscillates between *tradition and transcendence*. Building interfaces on metaphorical knowledge about the world (e.g., MONEY IS SUBSTANCE derived from “cashflow” or “a vital source of money”) emphasises tradition by focusing on existing mental models. However, image-schematic metaphors also lead to a high degree of transcendence and novelty by directing the interaction designer to think about the representation of abstract concepts in unconventional ways. Grounding the design process on image-schematic metaphors can, therefore, lead to unconventional ways of visual representations while at the same time maintaining a minimal form of tradition by aligning the user interface to the user's mental model. Since balancing tradition and transcendence has been described as one of the most fundamental challenges in design (Ehn, 1988; Madsen & Dalsgård, 2006), the promise of image-schematic metaphors to link both is important to the interaction design community.

On the dimension of *convergence and divergence*, image-schematic metaphors first steer the design process towards convergence by distilling image-schematic metaphors and thus mental models from large sets of the user's language. Only afterwards, image-schematic metaphors urge the interaction designer to diverge in the design process. Due to the possibility of creating multiple instantiations for each metaphor (see section 4.4.2), a multitude of diverse solutions and design ideas will be raised even though the majority of them might not be useful for later phases. However, in later phases of the design process, elements of convergence can be found again. For example, in the study of Antle et al. (2009), image-schematic metaphors served as a tool for mapping abstract concepts in the music domain to physical dimensions and tangible prototypes. Here, the metaphors narrowed down the number of prototypes by excluding prototype ideas that did not comply with the extracted metaphors. Thus, image-schematic metaphors can both drive the ideation process while at the same time channelling and constraining it based on their compliance.

Image-schematic metaphors can also improve the *structure of the design process* by employing image schemas as a meta-language for interaction design (Hurtienne, 2011). However, not many projects exist that utilised image-schematic metaphors in industry projects or standardised studies with multiple design teams (e.g. Hurtienne, 2011; Löffler, Hess, Maier et al., 2013). Thus, the method of image-schematic metaphors might provide structure and guidance during the design process, but empirical evidence for this claim is still sparse.

Finally and most importantly, image-schematic metaphors can be *sources for inspiration.* Image-schematic metaphors equip the interaction designer with a supplementary basis for stimulation and have been proposed to encourage innovative solutions and interfaces. Even though empirical data on this claim is sparse (Hurtienne, 2017a), Hurtienne, Klöckner et al. (2015) explicitly see the main advantages of this approach in the design process in "using metaphors for inspiration" (p. 7). Integrating image-schematic metaphors into existing design patterns can, therefore, be assumed to lead to the desired innovative user interfaces.

Besides image-schematic metaphors, similar sources for inspiration and the stimulation of creativity are available in the literature. The most similar design method to image-schematic metaphors is – without constraining the source domain to image schemas – applying general metaphors in the design process, an approach called *metaphorical design* (Blackwell, 2006; Madsen, 1994). Guidelines are available for finding useful interface guidelines such as "listen to how users interact with their system", "build on already existing metaphors" or "note metaphors already implicitly in the problem description (e.g., an existing link might stand for a pipe or a path)"

(Reeves, 1999, p.78). However, some researchers criticised metaphorical design as leading to non-scalable products and impractical or even senseless interactions (Blackwell, 2006; Cooper, 1995). The power of image-schematic metaphors turns out to be constraining the source domain to a set of pre-defined, frequently re-occurring and experience based image schemas (Hurtienne, 2011; M. Johnson, 1987). General interaction metaphors with too complex source domains often imply many entailments (Lakoff & Johnson, 1980) that can be confusing (e.g. ARGUMENT IS WAR with WAR as the source domain implies dozens of entailments), while restricted image schemas are more or less well defined and easy to instantiate into interactions and interface elements. Additionally, invariance hypothesis even states that image schemas (and not all facets of the concept of the source domain) are the important structure that is transferred to the target domain. Because complex metaphors are boiled down to specific image-schematic metaphors, the scalability is not in danger.

Besides metaphorical design as an approach for innovative solutions, Madsen and Dalsgård (2006) proposed another method for designing for innovation and creativity that is at least partially comparable to image-schematic metaphors. By combining so-called *Technology Cards* and *Domain Cards* (created during domain studies and technology studies, similar to user research and technology screening) in Inspiration Card Workshops, stakeholders from design-related or -unrelated disciplines are brought together in the design process. The approach of using single cards for inspiration is comparable to metaphorical design as well as designing with image-schematic metaphors: all three methods rely on a set of sources of inspiration that can stimulate the interaction designer to come up with innovative ideas. However, Inspiration Card Workshops do not guide interaction designers on the interface level and are not concentrated on the user's mental models of abstract concepts, which image-schematic metaphors promise to do.

Taken together, image-schematic metaphors match the general dimensions that are necessary for designing for innovation. Importantly, image-schematic metaphors promise to present concrete guidance on the interface level while at the same time stimulating the interaction designer's way of thinking.

4.5.2 IS-M and Age-Inclusiveness

Approaches like User-Centered Design or accessibility guidelines help interaction designers to adapt HCI to the needs and abilities of older adults. However, these approaches often lead to building separate technologies for older and younger adults that can even increase the gap between generations and can stigmatise older adults (Gooberman-Hill & Ebrahim, 2007; Walsh & Callan, 2011). Therefore, age-inclusive design is defined as the new standard (Hurtienne et al., 2013; Peace, 2016). This section aims to make explicit how image-schematic metaphors can facilitate the design of age-inclusive interfaces that can be used regardless of the user's cognitive resources or prior knowledge.

Using the cognitive linguistic approach of image-schematic metaphors (Hurtienne, 2011; M. Johnson, 1987) allows to capture technology-independent mental models through analysis of discourse and to map these mental models to design solutions systematically. At this point, the distinction between technology-dependent and technology-independent prior knowledge is crucial. On the one hand, younger and older adults are likely to have different exposure to technology and less technological competence. On the other hand, both age groups can draw on a fundamentally similar exposure to the physical world, culture and language and thus share basic mental models about abstract concepts that are in theory independent of age. The assumption is that technology-dependent prior knowledge (e.g. competence in using Skype) is a less appropriate foundation for age-inclusive interaction design than technology-independent prior knowledge (in the form of image-schematic metaphors, e.g., SOCIAL GROUPS ARE CONTAINERS). The automatic triggering of prior knowledge in the form of image-schematic metaphors minimises cognitive demand during those parts of interaction that are based on image-schematic metaphors.

The second goal of applying image-schematic metaphors in this work is therefore to provide a method for designing age-inclusive interactions. Grounding interfaces on image-schematic metaphors might render the interaction less dependent on the user's cognitive resources and prior knowledge because they activate an elementary and age-independent form of prior knowledge subconsciously. Image-schematic metaphors can equip interaction designers with universal and age-independent basic mental models. Even though empirical data on the universality of image-schematic metaphors across different ages is sparse, one of the major promises of these metaphors is age-inclusive design (Hurtienne, 2017a).

Finally, image-schematic metaphors need to be compared to other approaches that also promise age-inclusive interaction design. For example, Reddy (2012) suggested recommendations for interaction design that are based on empirical findings. Nurgalieva et al. (2017) similarly report on a set of studies that aim at channelling design decisions into the direction of age-inclusive interaction. However, these guidelines are, on the one hand, mostly constrained to sensorimotor and not on cognitive or prior knowledge-related problems (Hawthorn, 2007) and, on the other hand, often not adopted by interaction designers because they constrain them in their creativity (Fisk et al., 2009; Nurgalieva et al., 2017). Image-schematic metaphors can address this problem by providing for a set of metaphors for each design case that can be implemented in several ways (Hurtienne, 2011, study 7).

To conclude, image-schematic metaphors promise to provide an interaction design approach that addresses both innovation as well as age-inclusiveness at the same time. Few previous studies already provide insights on the question whether this promise can be met. In terms of perceived innovation, prototypes that were developed with image-schematic metaphors are often perceived as innovative and new (Antle et al., 2009; Hurtienne, Klöckner et al., 2015; Hurtienne & Langdon, 2010; Winkler et al., 2016). Although these prototypes are usually perceived as unfamiliar (Hurtienne, 2011; Hurtienne & Langdon, 2010; Winkler et al., 2016), interactions that conform with image-schematic metaphors are often intuitive to use (Bakker et al., 2012; Hurtienne, Klöckner et al., 2015; Hurtienne & Langdon, 2010; Hurtienne et al., 2010; Winkler et al., 2016). Age-related differences in cognitive abilities and prior technological knowledge have been shown to strongly affect HCI (Blackler et al., 2012; Hanson, 2011; Langdon et al., 2010). Interaction that is based on image-schematic metaphors promise to be not affected by these age-related differences. In other words, image-schematic metaphors promise to lead to age-inclusive HCI regarding cognitive abilities and prior technological knowledge.

However, concrete evidence for innovative and age-inclusive interaction design with image-schematic metaphors is sparse. First, one critical assumption of age-inclusive interaction design with image-schematic metaphors is that the image-schematic metaphors are universal across different age-groups (Hurtienne, 2017a). No evidence has been brought forward supporting this assumption yet. Second, the processes of extracting and designing with image-schematic metaphors has been applied in various projects (Asikhia et al., 2015; Hurtienne, 2011; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013; Lund, 2003; Winkler et al., 2016). Still, the reported insights on methodological improvements were often limited to case-studies with only one project team and no baseline team (Bakker et al., 2012; Hurtienne, Klöckner et al., 2015; Hurtienne & Langdon, 2010; Lund, 2003; Winkler et al., 2016) or no quantitative evaluation of the resulting prototypes (Löffler, Hess, Maier et al., 2013; Löffler et al., 2014). Third, previous studies investigating prototypes with image-schematic metaphors did not evaluate those either against any baseline (Antle et al., 2009; Bakker et al., 2012; Hurtienne, Klöckner et al., 2015) or against unfair baselines (Hurtienne et al., 2010), which renders the interpretation of ratings on perceived innovation and performance data difficult. Additionally, only few studies directly focused on the impact of age-related differences (Hurtienne, Klöckner et al., 2015) and no study directly measured cognitive abilities and prior technological knowledge – despite the fact that these direct measurements are far more important for assessing age-inclusiveness than age alone (Blackler et al., 2012; Vines et al., 2015). These open questions will be addressed in this thesis.

4.6 Research Questions

The present work is situated on the border between basic and applied science (Stokes, 1997) and aims at understanding human nature, deriving recommendations and evaluating prototypes. Even though understanding, recommending and designing are often not well interwoven in HCI and design research, the integration of these three research methodologies can draw full potential of the interdisciplinary field of HCI (Zimmerman & Forlizzi, 2014).

Image-schematic metaphors have been integrated into interaction design previously in various domains. The application of image-schematic metaphors in interaction design may be useful for developing a) innovative and b) age-inclusive user interfaces that are usable by very diverse user groups with differences in cognitive abilities and prior technological knowledge. Because image-schematic metaphors reveal fundamental and universal mental models of abstract concepts, grounding user interfaces on them should directly exploit the prior knowledge that can be found in younger as well as in older adults. Moreover, even though these interfaces seek to build on prior knowledge and experiences of the user, each image-schematic metaphor provides a certain flexibility in its concrete instantiations in the interface and might even serve as a source of inspiration. By this, designing with image-schematic metaphors could facilitate creative and innovative solutions that work for different age groups.

However, research needs to provide empirical evidence for some theoretical assumptions and progress the methodology before investigating the promises of image-schematic metaphors for innovative and age-inclusive interaction design.

R1 Do younger and older adults use the same image-schematic metaphors in natural language and thus substantially overlap in their mental models of abstract concepts?

Grounding interaction design on universal image-schematic metaphors and thus on universal mental models should lead to age-inclusive HCI. Image-schematic metaphors are an elementary form of technology-independent prior knowledge, and their application during interaction should not require mental effort. In fact, image schemas have been suggested and found in different languages, domains, age groups and modalities. A certain degree of universality of *image-schematic metaphors*, as the link between abstract concepts and specific image schemas, has also been shown, for example for different cultures (Kövecses, 2006; Löffler, 2017). However, the universality of image-schematic metaphors across different age groups is still an assumption (Hurtienne, 2017a) that must be validated before claiming the potential of age-inclusive interaction design with image-schematic metaphors. The first research question (*R1*) of this work, therefore, focuses on the overlap of younger and older adults in their mental models in terms of their usage of image-schematic metaphors.

R2 How applicable and creativity stimulating is the method of image-schematic metaphors in comparison to an industry standard method from the designer's perspective?

A high applicability and potential for stimulating innovative designs are important for the acceptance of the method from the designers' perspective. It is possible to integrating image-schematic metaphors into the User-Centred Design process and feedback obtained in previous studies is positive (Hurtienne, 2011; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013). However, methodological limitations like missing baselines (e.g. in form of a design team working without the method of image-schematic metaphors) or small sample sizes (Hurtienne, Klöckner et al., 2015;

Löffler, Hess, Maier et al., 2013) render the interpretation of these results challenging. Additionally, previous studies are often constrained to proof-of-concepts and case-studies, but comparisons to standard methods are rare (as one exception, see Löffler et al., 2014).

Also, applications of image-schematic metaphors in the design process often aim at translating requirements into interfaces that are intuitive to use (Hurtienne, 2011). Innovative interaction design has rarely been a focus but is one of the core promises of image-schematic metaphors (e.g. Hurtienne, Klöckner et al., 2015). Empirical evidence for this claim is still weak.

The second research question (*R2*) is interaction designers perceive the new method's applicability and potential for stimulating creativity compared to design method that represent the industry standard. It is expected that image-schematic metaphors can be readily integrated into the design process and provide an additional source of inspiration (Biskjaer et al., 2010) that supports the ideation phase (e.g. more ideas, creativity stimulating) compared to a conventional ideation phase (e.g. Holtzblatt & Beyer, 2017).

R3 Are user interfaces that explicitly follow image-schematic metaphors perceived as more innovative, and less dependent on cognitive abilities and prior technological knowledge compared to user interfaces that are not explicitly designed with image-schematic metaphors?

Finally, using image-schematic metaphors in interaction design promises innovative and age-inclusive interfaces (Hurtienne, 2017a). Image-schematic metaphors reveal basic mental models and, therefore, a form of technology-independent prior knowledge. This form of prior knowledge is almost automatically accessible during HCI and requires little cognitive effort. Together with the assumption that image-schematic metaphors are not only universal across cultures but also across age-groups (see also research question 1), image-schematic metaphors can provide an ideal basis for developing age-inclusive interfaces (see research question 2). However, the literature provides only little empirical data showing that prototypes based on elicited image-schematic metaphors can hold this promise. Even though predicted by theory, standardised experiments are still missing.

The third research question (R3), thus, aims at empirically evaluating user interfaces that were designed explicitly with image-schematic metaphors and comparing them with user interfaces that were designed using a standard methodology or represent the industry standard. Studies 5 and 6 focus on this research question and focus on different domains of interaction design.

Figure 4.2 summarises the three research questions and visualises the research framework underlying this thesis. Most importantly, the assumption of an overlap between younger and older adults in their basic mental models of abstract concepts instantiated in their natural language has not been subject to empirical investigation yet

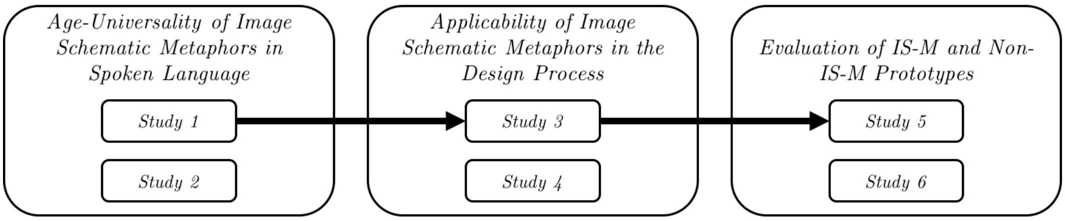


Figure 4.2: Research model of this thesis with the three research questions. For each research question, two empirical studies were conducted.

(neither in cognitive linguistics nor in HCI). Also, previous work more often than not did refrain from comparing developed image-schematic metaphor prototypes to a baseline, and the methodology of developing interactive prototypes can be seen critical. Core promises have often been investigated only on an explorative basis, and empirical data on both the theories assumptions and predictions is incomplete. Furthermore, various projects applied the method of integrating image-schematic metaphors in the design process, but methodological issues like missing baselines make the results difficult to generalise. These issues will have to receive attention before the question of whether image-schematic metaphors can support innovative and age-inclusive interaction design.

Chapter 5

Testing Age-Universality of Image-Schematic Metaphors

The first research question of this thesis is on the universality of image-schematic metaphors in natural language of different age groups. Under the assumption of Lakoff and Johnson (1980) that language can reveal underlying mental representations, the same image-schematic metaphors in the language of younger and older adults would indicate similarities in their mental representation of abstract concepts. These age-independent mental representations could then serve as a form of age-independent prior knowledge which could be exploited in age-inclusive HCI.

Due to their grounding in fundamental experiences, image-schematic metaphors should be universal across cultures and ages. Importantly, exploiting image-schematic metaphors as age-independent models for age-inclusive interaction design assumes that the same image-schematic metaphors should occur in natural language regardless of the speaker's age. In other words, the principal prerequisite for grounding age-inclusive interaction design in image-schematic metaphors is their universal occurrence in natural language. Many studies using image-schematic metaphors as a universal representation of abstract concepts implicitly take this assumption for granted (e.g. for children Antle et al., 2009; Hurtienne, 2017a; Hurtienne, Klöckner et al., 2015; Hurtienne et al., 2010).

Unlike cultural differences where the universal use of many image-schematic metaphors has been shown in natural language (Kövecses, 2006; Löffler et al., 2014), age differences have not been subject to systematic comparisons yet. Even though theory predicts the same image schemas in the language of younger and older adults (M. Johnson, 1987), it is an open question whether different age groups, in fact, share the same image-schematic *metaphors* (as the theory on primary metaphors would predict). The investigation of age-differences in image-schematic metaphors in natural language is crucial since younger and older adults differ considerably in their language (D. Burke & Shafto, 2008). For example, older adults know a more extensive variety of words (Kemper & Sumner, 2001) but have more difficulties in word production than recognition (D. Burke & Shafto, 2008; Kemper, 2012; Thornton & Light, 2006). On the other hand, younger adults usually possess more substantial prior technological knowledge over various domains which might lead to considerable differences in the word corpora used by younger and older adults (D. Burke & Shafto, 2008). Since younger adults are usually more experienced with technology-related domains (Czaja et al., 2006; Vorrink et al., 2017), they possibly use more, other, or just more specific words than older adults in this domain. Still, even though the natural language of younger and older adults differs on a word level, the underlying mental representations and semantic knowledge often do not (Salthouse, 2010) which should lead to the same image-schematic metaphors. Only then image-schematic metaphors can be used as a basis for age-inclusive interaction design (Hurtienne, 2017a).[1]

In the first two studies, the occurrence of image-schematic metaphors in natural language was used to reveal the participant's underlying mental model of abstract concepts. Even though other approaches exist (see also chapter 4, Hurtienne, 2017a; Hurtienne et al., 2010; Löffler et al., 2016), natural language is the most feasible way to elicit domain-specific image-schematic metaphors (Löffler, Hess, Maier et al., 2013). The aim of studies 1 and 2 is a comparison between younger and older adults regarding image-schematic metaphors in their natural language. Additionally, the extraction process of image-schematic metaphors from natural language has been reported as very time-consuming (Asikhia et al., 2015; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013). Thus, studies 1 and 2 provide insights into possible approaches to shorten the extraction process.

5.1 Study 1: Describing Abstract Concepts

To provide an estimation for the overlap of image-schematic metaphors in the natural language of different age groups, study 1 followed a standardised procedure. Contrary to free conversations, in which the language of the interviewer could bias word production, participants described a set of abstract concepts in their own words. These descriptions served as the basis for comparing the occurrence of image-schematic metaphors in the natural language and the mental models of younger and older adults.

Two domains were chosen for this study: *banking* and *everyday life*. The variety of abstract concepts from the domain of banking and everyday life is challenging for interaction design (Merdenyan, Kocyigit, Bidar, Cikrikcili & Salman, 2014; Vines et al., 2011, 2012). Both domains contain abstract concepts that usually possess no physical correlate and are difficult to map to physical dimensions (e.g., interest rate, friendship). Banking provides an example of a domain that changed drastically due to technological advances in the last years. Vines et al. (2012) argued that new services and customer interfaces in the banking sector like Near Field Communication (NFC)

[1] Still, when image-schematic metaphors are not age-inclusive, interaction designers could integrate both in the user interface at the same time to achieve age-inclusiveness.

and user interfaces that "allow people to transfer funds at the swipe of a finger" (p. 1169) are less appropriate for older adults. Prior technological experience of the domain of banking was expected to be significantly different for younger than for older adults. If prior technological experience influenced the underlying mental representations, the analysis of the natural language of younger adults should reveal different image-schematic metaphors compared to older adults. On the other hand, the influence of age-related differences in prior technological knowledge on conceptualisation of everyday life (e.g., with abstract concepts such as friendship, learning, remembering) was expected to be smaller. Even though modern technology (like Tinder and social networks) has been argued to influence social life (Turkle, 2017), core concepts of social life (like "social connectedness") should remain similar across different communication channels (Chayko, 2014). Taken together, the following two hypotheses were tested:

H1 Younger and older adults differ in their technology-dependent prior experience.

H2 Younger and older adults do not differ in their technology-*independent* prior experience of banking and everyday life and show an overlap in their use of image-schematic metaphors in natural language.

5.1.1 Method

Participants

41 German native speakers participated in the study. 21 of them were younger than 30 years (M = 21.7, SD = 2.3, min: 18, max: 27) and 20 older than 50 years (M = 68.4, SD = 12.7, min: 50, max: 86). In the group of younger adults, 12 were female and in the group of older adults 8. Younger adults were recruited via a local student panel. They received either course credit or 10 Euro for their participation. Older adults were recruited via direct contact, in healthcare facilities, and senior residences and all received 10 Euro. None of the participants was diagnosed severe cognitive decline (e.g., dementia) or showed age-untypical visual or auditive impairments. All younger adults had at least received Abitur, and only one had a university degree (bachelor). Older adults varied to a more considerable extent in their education ("none": one participant, "Mittlere Reife": 3, "Abitur": 12, "University degree": 3, "Ph.D": 1).

Procedure

The study was conducted in one-on-one sessions either in the university's laboratory or – when more accessible for older adults – in their health-care facility or at home. In the first part, participants completed a questionnaire on demographics and prior technological knowledge. Due to their age, three older participants were physically not able to complete all questionnaires by themselves. In these cases, the interviewer read out loud the questions and wrote down the answers.

The second part of the study focused on the collection of the natural language of participants. Using a standardized instruction, each participant described a set of 26 abstract concepts (see section 5.1.1 "Material"). For this, participants were instructed to describe the abstract concepts shortly in their own words and only in a few sentences. To reduce biases, the interviewer did not provide any further help in describing each abstract concept. For each concept, participants answered the question "What is [abstract concept]". To standardise the instruction for the description of the single concepts, a list of the 26 abstract concepts was used during the interview and abstract concepts were not randomised (see appendix A.1.1). This part of the interviews was audio-taped and later transcribed for the following image schema and image-schematic metaphors analysis.

Depending on individuals differences, the entire procedure took between 45 and 120 minutes. For example, older adults usually had more questions regarding the procedure and were not as familiar with questionnaires as younger adults, increasing the time for completing the questionnaires.

Material

To generate a set of common abstract concepts in the main study, a pre-study was conducted. Six participants (age: 23, 27, 32, 40, 56, 66 years) completed open-structured interviews (duration between 45 and 90 minutes) on the participants' usage of banking and everyday life. These domains were regarded as containing a variety of possible abstract concepts. The final 26 abstract concepts occurred most frequently during the interviews (see Table 5.1). 13 abstract concepts originated from the domain of banking and 13 from the domain everyday life. Note that no image-schematic metaphors were extracted in the pre-study. Additionally, the abstract concept of energy was inspired for the main study by Löffler et al. (2014) and had not been part of the pre-study.

Prior technological experience was measured on the two dimensions exposure and competence (Hurtienne et al., 2013). Both dimensions were measured on three different levels of specificity (low, medium, high). Data on the participants' prior technological knowledge was collected that covered technological exposure and competence in general (low and medium specificity: relevant for concepts for everyday life) as well as prior knowledge specific to the domain "Online-Banking". This approach led to the following six facets of prior technological knowledge:

exposure - low specificity Participants reported how often they had used each of 15 different technologies in their life (e.g. a ticket vending machine, smartphone, personal computer; from 'never' to 'regularly'; range of sum score: 15 to 75).

exposure - medium specificity Participants reported how often they had per-

Table 5.1: Abstract concepts used in the study (German [English]). For the detailed instruction, see appendix A.1.1

Banking	Everyday life
GELD [MONEY]	ERINNERUNG [REMEMBERING]
BANK [BANK]	BESTELLUNG [ORDER]
ÜBERWEISUNG [TRANSACTION]	RESERVIERUNG [RESERVATION]
ÜBERWEISUNG - HANDLUNG [TRANSACTION - ACTION]	EMPFEHLUNG [RECOMMENDATION]
ZINSEN [INTEREST]	TELEFONAT [CALL]
ZINSEN - GRUND [INTEREST - REASON]	BEGRÜSSUNG [WELCOME]
SPAREN [SAVING]	FREUNDSCHAFT [FRIENDSHIP]
DAUERAUFTRAG [STANDING ORDER]	ABENTEUER [ADVENTURE]
DAUERAUFTRAG - HANDLUNG [STANDING ORDER - ACTION]	UNTERHALTUNG [RANDOM CONVERSATION]
KONTO [BANK ACCOUNT]	ENERGIE [ENERGY]
NEGATIVER KONTOSTAND [NEGATIVE BALANCE]	ERKENNTNIS [INSIGHT]
MEHRERE KONTEN [MULTIPLE BANK ACCOUNTS]	GESPRÄCH [DIRECTED CONVERSATION]
KUNDENBERATER [CUSTOMER ADVISER]	LERNEN [LEARNING]

formed each of 20 different tasks on a screen-based technology (e.g., listening to music, social media, sending e-mails; 'never' to 'daily'; range of sum score: 0 to 100).

exposure - high specificity Participants reported in a checklist which of 21 functions common in *online* banking they had ever used in their lives (e.g., change personal information such as the address, perform a transaction, check finance status; range of sum score: 0 to 21).

competence - low specificity Participants completed a questionnaire on technology affinity to electrical devices (TA-EG; Karrer et al., 2009). This questionnaire distinguishes between four primary dimensions perceived technology competence, interest in technology, and positive and negative attitudes towards technology (range of medium score: 1 to 5).

competence - medium specificity Participants completed the Computer Literacy Scale (CLS; Sengpiel & Dittberner, 2008), a performance test for computer icons (range of sum score: 0 to 25).

competence - high specificity Participants subjectively rated their competence for online-banking on a 7-Likert scale (range: -3 to +3 with a neutral "0").

Image schemas extraction process

The descriptions of the 26 abstract concepts by 41 participants allowed for the collection of 7 hours, 10 minutes, and 24 seconds of natural language. 36 (3.4%) of the descriptions were missing data (younger adults: 6 out of 546 descriptions; older adults: 30 out of 520 descriptions), that could be tracked down to technical problems or participants' difficulties to describe a single abstract concept (e.g. "I don't want to talk about friendship.", "A bank is something bad, I cannot say more"). 1025 of all 1066 possible descriptions were included in the further data analysis and analysed using Excel 2015 and SPSS version 24. All descriptions were transcribed on a word-to-word basis (in contrast to only summarizing their content, but leaving out pauses and filler words such as "ähm" [uhm]). Younger adults explained each single concept on average with descriptively but not significantly fewer words (M = 27.9, SD = 15.7) than older adults (M = 30.3, SD = 19.4), a phenomenon that has been described earlier (Arbuckle & Gold, 1993) as an increasing verbosity in age. Significant word amount differences between both age groups occurred for the two concept of FRIENDSHIP (younger adults: M = 21.0, SD = 11.4; older adults: M = 36.6, SD = 25.7), $t(26) = 2.51$, $p = .019$, $d = .79$, and ADVENTURE (younger adults: M = 19.5, SD = 10.7; older adults: M = 33.7, SD = 26.1), $t(26) = 2.11$, $p = .046$, $d = .71$.

Due to the large dataset and the limited resources of trained extractors, the author of this work manually extracted image schemas and image-schematic metaphors from all 1025 descriptions. Text analysis in overall profits from word lists that map keywords to investigated concepts which has been shown to increase reliability of analysis methods similar to image-schematic metaphor extraction (Doucet & Jehn, 1997; Gephart, 1993). The extraction of image schemas and image-schematic metaphors from transcripts followed a set of primary keywords (Löffler, Hess, Hurtienne et al., 2013) to map a small set of German words directly to single image schemas. However, this set of keywords did not cover all image schemas, and its keywords were not sufficient for a complete extraction of all descriptions. Thus, the list was used as a basis for each abstract concept. Especially for metaphor analysis, the extraction process also benefits from iteratively developed lists of words or phrases for signalling metaphorical meaning (Gibson & Zellmer-Bruhn, 2001).

Every time a keyword, syllable,[2] or formulation led to the extraction of an image-schematic metaphor, the extractor checked whether occurrences of the same phrases had also led to the extraction of image-schematic metaphors in the other participants' descriptions of the same concept. By this, increased internal consistency and reliability of the extraction were expected.

To avoid sequence effects of the extraction process, each of the 26 concepts was analysed separately after the transcription of all interviews. Also, the extraction order of participants was randomised for each concept, and it was not clear to the extractor if the current description stemmed from a younger or older participant. Only metaphorically meant phrases led to the extraction of image-schematic metaphors. When single words were ambiguous and did not contain keywords (e.g., "Geld verschieben"[push money] for COMPULSION), the German Duden (Duden, 2017) was consulted to validate the metaphorical meaning of a word in this context. Per each abstract concept, the extraction process focused only on image-schematic metaphors in the form of

[2] Following the methodology of previous work (Hurtienne, Klöckner et al., 2015; M. Johnson, 1987; Löffler, Hess, Maier et al., 2013) we also analyzed single pre- and suffixes for image schemas (e.g., "an-schalten"[turn on] for CONTACT).

Table 5.2: Prior technological knowledge. ELS: Exposure low specificity; EMS: Exposure medium specificity; EHS: Exposure high specificity; CLS: Competency low specificity; CMS: Competency medium specificity; CHS: Competency high specificity. TA-EG: Technology competence questionnaire (Karrer et al., 2009). CLS-B: Computer literacy scale - performance test (Sengpiel & Dittberner, 2008).

	younger Adults	older Adults	t-test (adjusted df)	p	Cohen's d
Prior technological knowledge					
exposure hardware (ELS)	43.7 [6.2]	28.4 [13.7]	$t(22.6) = 4.36$	<.001*	1.46
exposure digital tasks (EMS)	60.7 [9.0]	36.3 [17.5]	$t(19.1) = 4.97$	<.001*	1.86
exposure online banking tasks (EHS)	6.0 [3.4]	3.4 [5.8]	$t(26.4) = 1.73$	.11	.55
TA-EG - competency (CLS)	3.7 [0.7]	2.8 [0.7]	$t(38.0) = 4.50$	<.001*	1.43
CLS-B performance test (CMS)	25.3 [0.9]	11.6 [9.9]	$t(13.2) = 5.20$	<.001*	2.24
Competency online banking (CHS)	0.5 [1.6]	-1.4 [2.2]	$t(36) = 3.10$	.003*	1.03
Others					
TAEG - enthusiasm	3.4 [0.5]	2.8 [0.6]	$t(38) = 6.6$	.004*	1.09
TAEG - positive attitudes	3.6 [0.5]	3.4 [0.7]	$t(38) = 1.4$	.17	.33
TAEG - negative attitudes	3.0 [0.9]	2.8 [0.8]	$t(38) = 0.7$	.50	.24

[CURRENT CONCEPT] is [IMAGE SCHEMA], for example, SAVING MONEY IS PERIPHERY from "Beim Sparen lege ich Geld beiseite" [When saving money, I put money aside]. Importantly, only image-schematic metaphors for the pre-defined 26 abstract concepts were extracted. Thus, speech structuring comments (e.g. "as I already said *before*") and abstract concepts that were irrelevant for the project (i.e. all abstract concepts except the 26 pre-defined ones; e.g. "*in* summer) were ignored in the extraction process. The final dataset consisted of 1295 instances of image-schematic metaphors for all 26 concepts (younger adults: 749; older adults: 546). In total, 137 different image-schematic metaphors with at least two instances were extracted.

5.1.2 Results

Age-differences in prior technological knowledge

Table 5.2 summarises the prior technological knowledge of both age groups. Younger and older adults differed significantly in five out of the six measured levels of technology-dependent prior knowledge as expected. Younger adults reported significantly higher values on most scales. The difference between both age groups was not significant only for exposure to technology with high specificity for the domain of online banking (number of online banking functions used previously). On the TA-EG-scales regarding positive and negative attitudes towards technology, younger and older adults did not differ significantly. These results show that the two age groups in the sample differed significantly in their technology-dependent prior knowledge both in general terms as well as specific to the domain of online banking.

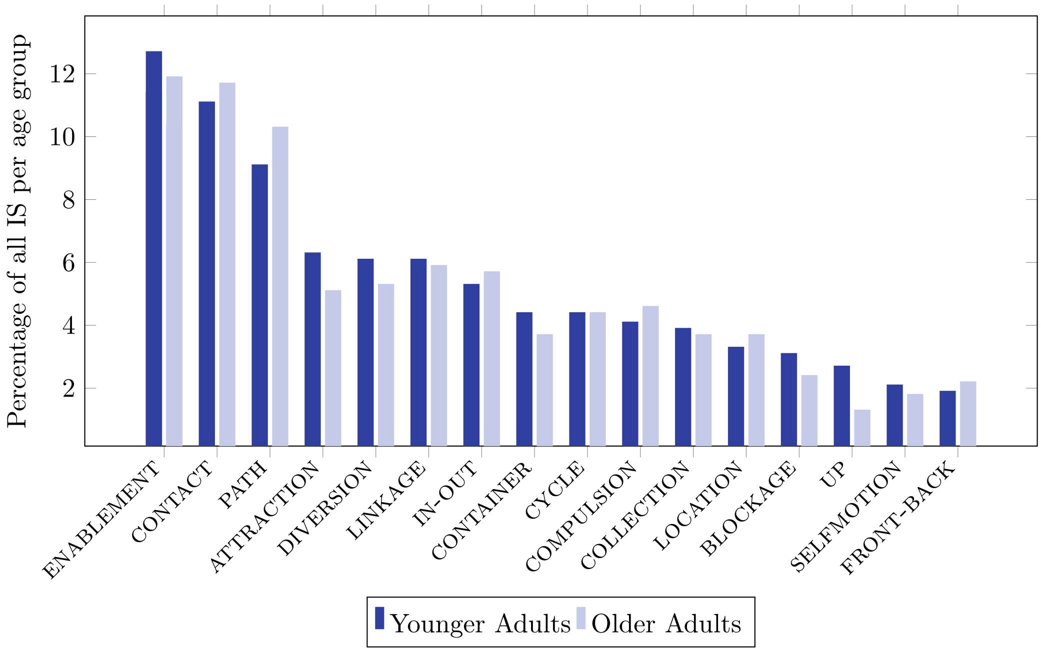


Figure 5.1: Most frequent image schemas (IS) used by younger and older adults.

Image schemas per age group

Over the entire corpus of analysed natural language from all descriptions, image schemas from the categories FORCE and SPACE were used most often by both age groups, followed by CONTAINMENT and MULTIPLICITY (see figure 5.3). Figure 5.1 shows the most frequent single image schemas per age group over all descriptions. The two age groups show a similar pattern in the distribution of image schemas in natural language. Even when considering only image schemas with at least one instantiation over all participants (39 image schemas), the two age groups highly correlate in their distribution of image schemas, $r(38) = .983$, $p < .001$. Comparing the ranks of image schemas between both age groups did not reveal any significant differences, U =724.5, $p = .719$. Thus, no overall effect of age was found on the distribution of image schemas.

Table 5.3: Frequency of image schemas categories per age group.

Category	Younger Adults	Older Adults	Overall
FORCE	36.2%	33.5%	34.9%
SPACE	30.7%	31.8%	31.2%
CONTAINMENT	13.8%	15.7%	14.6%
MULTIPLICITY	14.8%	13.3%	14.2%
PROCESS	4.4%	4.6%	4.5%
ATTRIBUTE	0.1%	1.1%	0.5%

Image-schematic metaphors per age group

If both age groups' underlying mental representations of the 26 concepts were identical, Conceptual Metaphor Theory would predict the same image-schematic metaphors in both age group's natural language. Because the focus was on natural language, a complete standardisation of the answers was not feasible, and a high variance between the answers and therefore topics and descriptions was expected. For example, when participants described the concept of money, only MONEY IS ENABLEMENT was included even when the participant randomly also talked about the concept of saving money (thus, a metaphor like SAVINGS ARE BACK was discarded).

It was possible that single descriptions did not contain any image-schematic metaphor. In sum, descriptions of older adults revealed less metaphorical language. For example, one participant described the concept of a bank with "A bank is something evil, I cannot say more". Over all 26 concepts, each description by a younger adult contained 1.40 instances of image-schematic metaphors (SD = .67), but each description by an older adults contained 1.11 instances of image-schematic metaphors (SD = .60; difference not significant, $t(50) = 1.6$, $p = .11$).

The descriptions of younger adults contained more instances of image-schematic metaphors. However, similar image-schematic metaphors were found in both age groups. Since Löffler, Hess, Hurtienne et al. (2013) suggested to consider even image-schematic metaphors with only two instances, it was refrained from comparing the proportion of each single image-schematic metaphor between age groups and, instead, focus on the finally extracted metaphors. Thus, the overlap was determined via the proportion of image-schematic metaphors (not the proportion o instances) that occurred in by both age groups.

For example, both age groups talked about a negative bank account by using only the two image-schematic metaphors NEGATIVE BANK ACCOUNT IS CONTAINER ("da versuche ich aus den roten Zahlen *raus*zukommen" [I try to get out of the red numbers]) and NEGATIVE BANK ACCOUNT IS BLOCKAGE ("dann kann ich *nichts* mehr abheben" [I cannot withdraw any money anymore]). In this case, the overlap between both age groups was 100% because both age groups used only metaphors that the other age group used as well. As another example, both age groups used SAVING MONEY IS BACK ("Geld zurücklegen" [put back money]), and SAVING MONEY IS BLOCKAGE ("wenn man sein Geld *nicht* ausgibt" [when I do not spend my money]). However, younger adults also used SAVING MONEY IS PERIPHERY ("Geld *beiseite* legen" [to put money aside]) and only older adults used SAVING MONEY IS CONTACT ("Geld *an*legen" [to invest money]). Here, the overlap for the concept saving money was only 50% (two overlapping concepts by four). Following Löffler, Hess, Hurtienne et al. (2013), image-schematic metaphors with only one instance per age group were discarded.

For reasons of simplification, Table 5.4 summarises this analysis for the top three image-schematic metaphors with the most instances per concept. Taken together, the overlap between both age groups is 74.8% (SD = 22.9%)[3], which represents a substantial agreement between groups (Landis & Koch, 1977). For the domain of online banking, the overlap between both age groups was 69.7% (SD = 21.2%) and for concepts related to everyday life 79.2% (SD = 24.6%). This difference between both domains, though, is not significant, $t(24) = 1.01$, $p = .32$. The overlap between both age groups differs between the single concepts. While for 10 out of 26 concepts (38.5%) exactly the same image-schematic metaphors were found in both age groups, nine concepts (34.6%) showed an overlap of at least 66%, and for seven concepts (26.9%) the overlap was 50% or less.

In general, no image-schematic metaphor was used by all 41 participants. The highest percentages of younger adults using the same image-schematic metaphors were for TRANSACTION IS PATH and STANDING ORDER IS CYCLE, both with 95.2%. In the group of older adults, 90.0% used STANDING ORDER IS CYCLE and 78.9% REMEMBER IS CONTACT. Very stable and often occurring image-schematic metaphors (i.e., more than 50% in both age groups) were TRANSACTION IS PATH, STANDING ORDER IS CYCLE, BANK ACCOUNT IS LOCATION, and REMEMBERING IS CONTACT. At least one age-independent image-schematic metaphor was always found for each concept. In 19 out of the 26 concepts (73.1%), the most frequent image-schematic metaphor was identical in both age groups. In 23 out of 26 concepts (88.5%), the most frequent image-schematic metaphor of each age group was also under the top 3 image-schematic metaphors of the other age group.

Cohen's Kappa score (Landis & Koch, 1977) adjusts the percentage of agreement between groups to chance based on the expected agreement. Still, the number of categories (i.e. image schemas that could have been used to talk about each abstract concept) was very high and expected agreement due to guessing was minimal.[4] Thus, the same thresholds as for the interpretation of kappa values served as the basis for the interpretation of the overall overlap percentage in this study. Following this classification, an overlap of more than 60% is substantial and of more than 80% (almost) perfect.

5.1.3 Discussion

In study 1, 41 participants described 26 abstract concepts from the domains of banking and everyday life using their own natural language. Additionally, prior technological knowledge was measured on six different levels (Hurtienne et al., 2013). The hypothesis was that – even though younger adults should report significantly less prior technological knowledge (H1) – younger and older adults should overlap

[3]The same analysis was also conducted with comparing all image-schematic metaphors per abstract concept (up to 8 image-schematic metaphors per concept with more than one instance in one age group). Here, the overall overlap was 77.2% (SD = 34.3%).

[4]Kappa is calculated via the formula $\kappa = \frac{p_o - p_e}{1 - p_e}$ with p_o as the observed agreement and p_e as the expected agreement by chance. Since the number of categories is very high, p_e can be regarded as very small justifying to handle the raw percentages as similar to kappa values.

Table 5.4: Most frequent image-schematic metaphors for each abstract concept per younger and older adults. Note that only image-schematic metaphors with at least two instances per age group were considered. Because some descriptions did not reveal any instance of an image-schematic metaphor, the percentages do not add up to 100%. Also, some concepts only revealed less than three image-schematic metaphors per age group. RR: RESTRAINT-REMOVAL. *N: Number of abstract concept's descriptions available for the age group.*

Concept	Group	Rank 1 [%]	Rank 2 [%]	Rank 3 [%]	N	Overlap
Geld	Younger	ENABLEMENT 76.2	MATCHING 14.3	-	21	1.00
[Money]	Older	ENABLEMENT 21.1	MATCHING 10.5	-	19	
Bank	Younger	LOCATION 38.1	CONTAINER 33.3	SURFACE 9.5	21	0.33
[Bank]	Older	LOCATION 42.1	-	-	19	
Überweisung	Younger	PATH 95.2	COMPULSION 61.9	UP / CONTACT 38.1	21	0.75
[Transaction]	Older	PATH 73.7	CONTACT 31.6	COMPULSION 26.3	19	
Überweisung - Handlung	Younger	DIVERSION 57.9	COLLECTION 52.6	FULL / CONTACT 47.4	19	0.40
[Transaction - Action]	Older	IN 55.0	COLLECTION 55.0	FULL 50.0	20	
Zinsen	Younger	UP 47.6	PART 28.6	CYCLE 28.6	21	0.75
[Interest]	Older	UP 15.0	CYCLE 15.0	PART / SCALE 10.0	20	
Zinsen - Grund	Younger	ENABLEMENT 38.1	ATTRACTION 28.6	BLOCKAGE 9.5	21	0.67
[Interest - Reason]	Older	ENABLEMENT 18.8	BLOCKAGE 12.5	-	16	
Sparen	Younger	BLOCKAGE 55.0	PERIPHERY 35.0	BACK 25.0	20	0.50
[Saving Money]	Older	BLOCKAGE 21.1	CONTACT 21.1	BACK 21.1	19	
Dauerauftrag	Younger	CYCLE 95.2	SELF-MOTION 33.3	PATH / CONTACT 23.8	21	0.75
[Standing Order]	Older	CYCLE 90.0	SELF-MOTION 30.0	CONTACT 30.0	20	
Dauerauftrag - Handlung	Younger	DIVERSION 55.0	IN 40.0	CONTACT / COLLECTION 35.0	20	0.75
[Standing Order - Action]	Older	DIVERSION 21.1	IN 21.1	COLLECTION 21.1	16	
Konto	Younger	ENABLEMENT 81.0	LOCATION 76.2	CONTAINER 57.1	21	0.50
[Bank Account]	Older	LOCATION 63.2	ENABLEMENT 52.6	SURFACE 36.8	19	
Negativer Kontostand	Younger	CONTAINER 40.0	BLOCKAGE 10.0	-	20	1.00
[Negative Bank Account]	Older	CONTAINER 26.3	BLOCKAGE 10.5	-	19	
Mehrere Konten	Younger	COLLECTION 57.1	SPLITTING 28.6	PART 23.8	21	1.00
[Multiple Bank Accounts]	Older	SPLITTING 26.3	COLLECTION 21.1	PART 15.8	19	
Kundenberater	Younger	ENABLEMENT 90.5	LINKAGE 19.0	COMPULSION 9.5	21	0.67
[Bank Customer Adviser]	Older	ENABLEMENT 36.8	COMPULSION 10.5	-	19	
Erinnerung	Younger	CONTACT 52.4	IN 38.1	ENABLEMENT 14.3	21	1.00
[Remembering]	Older	CONTACT 78.9	ENABLEMENT 21.1	IN 15.8	19	
Bestellung	Younger	ATTRACTION 66.7	CONTACT 28.6	PATH 14.3	21	1.00
[Order]	Older	ATTRACTION 35.0	CONTACT 15.0	PATH 10.0	20	
Reservierung	Younger	ATTRACTION 61.9	BLOCKAGE 28.6	FRONT / BACK 14.3	21	0.40
[Reservation]	Older	ATTRACTION 42.1	ENABLEMENT 26.3	BACK 21.1	19	
Empfehlung	Younger	PATH 47.6	DIVERSION 23.8	CONTACT 19.0	21	0.75
[Recommendation]	Older	DIVERSION 22.2	ENABLEMENT 16.7	PATH 16.7	18	
Telefonat	Younger	LINKAGE 47.6	CONTACT 28.6	ENABLEMENT 23.8	21	1.00
[Phone Call]	Older	LINKAGE 44.4	CONTACT 44.4	ENABLEMENT 11.1	18	
Begrüßung	Younger	DIVERSION 47.6	CONTACT 9.5	-	21	1.00
[Welcome]	Older	DIVERSION 38.9	CONTACT 16.7	-	18	
Freundschaft	Younger	LINKAGE 61.9	ATTRACTION 33.3	SPLITTING 19.0	21	0.40
[Friendship]	Older	ENABLEMENT 50.0	LINKAGE 35.0	ATTRACTION / CONTACT 15.0	20	
Abenteuer	Younger	IN 9.5	ATTRACTION 9.5	RR 9.5	21	0.75
[Adventure]	Older	IN 27.8	OUT 27.8	ATTRACTION / RR 16.7	18	
Unterhaltung	Younger	LINKAGE 42.9	DIVERSION 14.3	ENABLEMENT 9.5	21	1.00
[Conversation]	Older	LINKAGE 44.9	DIVERSION 16.7	ENABLEMENT 11.1	18	
Energie	Younger	COMPULSION 47.6	CONTACT 23.8	ENABLEMENT 14.3	21	1.00
[Energy]	Older	COMPULSION 47.4	ENABLEMENT 31.6	CONTACT 10.5	19	
Erkenntnis	Younger	OUT 28.6	IN 23.8	PATH / DIVERSION 14.3	21	0.75
[Insight]	Older	IN 31.6	OUT 26.3	PATH 15.8	19	
Gespräch	Younger	CONTAINER 26.3	LINKAGE 21.1	PATH 21.1	19	1.00
[Directed Conversation]	Older	CONTAINER 38.9	LINKAGE	PATH 22.2	18	
Lernen	Younger	CONTACT 45.0	IN 25.0	ENABLEMENT 25.0	20	0.50
[Learning]	Older	PATH 26.3	CONTACT 21.1	ENABLEMENT 15.8	19	
						0.748

in their technology-independent prior knowledge revealed through image-schematic metaphors analysis of their natural language (H2).

As predicted, older adults showed significantly less prior technological knowledge on five of six levels. The sample of younger adults reported to perform digital tasks more regularly, was exposed to more digital devices, performed better on an objective computer literacy task and subjectively rated their competence with online banking and technology in general higher compared to the group of older adults. On one level of technology-dependent prior knowledge - exposure to online banking tasks - only a descriptive difference could be found. The not significant difference on this level could have been caused by a bottom effect meaning that also younger adults had not used more than a few essential functions of online banking (as the questionnaire only asked for functions they had ever used). Thus the difference between both groups might have been underestimated by this measurement. However, based on the other five levels, it can be assumed that both age groups significantly differed in their prior technological knowledge, which supports H1.

Both age groups show a similar distribution of image schemas in their language. The distribution of image schemas does not differ significantly between both age groups, supporting the claim that image schemas are used universally (M. Johnson, 1987), also across individuals from different generations. However, more importantly, results indicate a substantial overlap of image-schematic metaphors between both age groups of 74.8% over all concepts. Even though the same image-schematic metaphors have been reported in different cultures (Kövecses, 2006; Löffler, 2017), this study for the first time shows that a substantial part of image-schematic metaphors is shared by younger and older adults stemming from different environments and technological backgrounds. Even though the overlap is highly dependent on the specific abstract concept, H2 is supported at least partially as well.

Following the link between image-schematic metaphors and mental models (Hurtienne, 2011; M. Johnson, 1987), younger and older adults seem to represent most concepts similarly regarding the most frequent image-schematic metaphors. However, the in-depth inspection of the image-schematic metaphors revealed interesting differences between age groups. For example, FRIENDSHIP is represented by older adults as a form of help and enablement, while younger adults emphasise its character as sharing and distributing information and experiences amongst friends. For this concept, the overlap between both age groups was 40%. While both age groups adopted FRIENDSHIP IS LINKAGE ("eine Freundschaft basiert auf viel Spass *mit*einander" [a friendship is based on fun *with* each other]) and FRIENDSHIP IS ATTRACTION ("ein Freund ist jemand, zu dem ich mich *hingezogen* fühle" [a friend is someone I am *attracted* to]), only older adults used FRIENDSHIP IS ENABLEMENT ("Freunde sind Leute, auf die ich mich verlassen *kann*" [friend are people I *can* rely on]). This finding might show that, especially in social concepts, different life durations and experiences of younger and older adults can bias their mental models and thus image-schematic metaphors.

Also, SAVING MONEY is represented by both age groups as a BLOCKAGE. Still, while both age groups also use the image schema BACK to talk about the concept of saving money, only younger adults use the image schema PERIPHERY in their natural language for this concept. Another example of different image-schematic metaphors between both age groups is the actions involved in carrying out a transaction. Younger adults talked more about different ways of carrying out a new transaction ("ich nutze entweder meine App *oder* gehe an den Automaten" [I either use my app or the ATM]), resulting in DOING A TRANSACTION IS DIVERSION. Older adults more focused on information that has to be inserted in banking, resulting in DOING A TRANSACTION IS IN ("da gebe ich alle meine Daten *ein*" [I insert all my information]).

Another example is RESERVATION, where (besides a shared image-schematic metaphor RESERVATION IS ATTRACTION) younger adults emphasise its character of BLOCKAGE from something for someone else. Older adults, on the other hand, focus on its aspect of ENABLEMENT for their later visit or product shipment. These examples show two things. First, when analysing natural language from younger and older adults and extracting image-schematic metaphors for abstract concepts, a satisfying general overlap can be expected. Second, however, minor differences will occur. Grounding interaction decisions only on image-schematic metaphors extracted from one age group will ignore these differences, which might be highly relevant for interaction design. The reason for these differences could be based on different experiences and stages of life (e.g. friendship conceptually but as well practically means something different for younger than for older people). Differences might also stem from different levels of expertise and technologically well-versed younger adults possibly described concepts on a higher level and chunked information more than older adults. Still, the question remains whether these differences are relevant for interaction design or not. Taken together, conceptual differences cannot be ignored. Still, incorporating image-schematic metaphors from different age groups into one user interface might be the best solution when no age-universal metaphor can be found for a concept.

Several limitations of this study must be taken into account when interpreting these results. First, only prior technological knowledge was measured even though both age groups should also differ in other aspects like cognitive abilities. In this study, the focus was on prior knowledge because a more extensive word variety in the domain specific language presumably influenced the choice of words more than cognitive resources. It was expected that lower cognitive capabilities would not influence the content of verbal expression of older adults as strongly as prior exposition to domain-related language. Even though older adults experience more difficulties in word production than younger adults, semantic representation and involved cognitive processes are maintained in old age (Botwinick, 1977; Wingfield, Lindfield & Harold, 2000). In general, lexical diversity increases as people get older (D. Burke & Shafto, 2008), and age-related problems of word production often originate from phonological and orthographical issues (Mortensen, Meyer & Humphreys, 2006). Less prior technological knowledge, on the other hand, was expected to strongly influence the corpus of words and concepts that was known to each participant and bias the image-schematic metaphors in natural language. Also, one question on prior technological knowledge asked which of a set technology participants had already used (and how often) *in their life*. Even though this question could focus instead on, for example, only the last twelve months, the time frame of this question was deliberately extended to the lifetime to provide a basic estimation on technology use not only in the recent past, but also on the time when older adults still had access to more technology at the workplace. Second, the extraction process was conducted by only one person, and systematic errors and biases in the extraction process could have influenced the final dataset of image-schematic metaphors. Several precautions (keywords, iterative analysis, randomised order, supervision by a second expert) presumably increased the overall reliability and validity, but replications of this study are desirable.

Using natural language to reveal underlying mental representations, study 1 shows that younger and older adults, in fact, share a substantial percentage of image-schematic metaphors. This finding is promising since it supports one central claim of Hurtienne (2011) and M. Johnson (1987): that age-effects (most prominently different prior experiences) on the use of image-schematic metaphors are limited. In sum, using image-schematic metaphors can be therefore regarded as a good candidate for age-inclusive interaction design. To avoid biasing the participants' word production, no interaction with the interviewer was allowed than went further than stating the current abstract concept that had to be described. Still, one could argue that this form of natural language is not in fact 'natural'. Thus, the next study focused on natural language collected in a more natural context.

5.2 Study 2: Contextual Inquiries

Study 1 revealed a substantial overlap of technology-independent prior knowledge in the form of image-schematic metaphors in the natural language of younger and older adults. Still, the methodology was standardised, and the large sample size is not realistic for industry-projects. Study 2, therefore, addresses the open question of age differences in the extraction of image-schematic metaphors when efficiency is crucial, and time and participants are constrained.

One possibility to access image-schematic metaphors efficiently is to exploit natural language occuring at early stages of the User-Centred Design process. One of the standard methods in this area is the Contextual Interview during Contextual Design (Holtzblatt & Beyer, 2017; Holtzblatt et al., 2005). The domain of this study was inspired by Knobel et al. (2012) who reported on the idea of creating a social experience for people distributed over different cars by providing means to communicate (e.g. call) and share entertaining content (e.g. music recommendations). The content of the Contextual Interviews thus embraced a wide range of abstract concepts from social communication and digital entertainment between passengers of different cars. During a Contextual Interview, real users are observed and interviewed on-site while interacting with their current system and context. Contextual Interviews allow for the recording of the users' natural language while talking about project-relevant topics and concepts. Usually, the project team does not analyse this linguistic material, and recordings often only serve for documenting the high-level structure of the Contextual Interview. However, the recorded natural language can also serve as the basis for extracting image-schematic metaphors (Hurtienne, 2011; Hurtienne, Klöckner et al., 2015). Since the Contextual Interviews cover most issues relevant to the later design, the majority of project-relevant abstract concepts should emerge. However, age differences have not been the focus of previous research for this form of language basis. The investigation of these differences is a necessity for providing recommendations for extracting image-schematic metaphors in practice. The aim of study 2 is therefore twofold:

First, the aim is a replication of the findings of study 1 regarding the influence of the participants' age on resulting image-schematic metaphors. Importantly, the task of describing single words could have constrained participants in their word production. Especially when describing things (e.g., images or other stimuli), older adults have been reported to produce more utterances that are irrelevant for the description itself (Bortfeld, Leon, Bloom, Schober & Brennan, 2001; D. Burke & Shafto, 2008). Also, age-effects of word production have been shown to differ between standardised tests and more ecologically valid settings based on conversations (Vesneski, Schmitter-Edgecombe & Jones, 1998). Even though word production and underlying mental representations are probably independent (D. Burke & Shafto, 2008), different tasks for word production (standardised, conversational) still could bias the occurrences of keywords for the extraction of image-schematic metaphors. Thus, in study 2, image-schematic metaphors are extracted from natural language from conversations during Contextual Interviews with younger and older adults.

Second, the methodology of the extraction process of image-schematic metaphors must be improved. Study 1 was – from a practitioner's perspective – inefficient and would not be feasible in most real projects. In study 2, the extraction process of image-schematic metaphors is carried out under realistic conditions by trained, but not long-term experienced extractors. This process is prone to errors and ambiguities (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013), even in very supervised teams, where extractors extracted image-schematic metaphors from given text-fragments of Contextual Interviews (Hurtienne, 2011). Still, these studies report a moderate to substantial agreement between different extractors. Study 2 investigates how reliable the extraction process of image-schematic metaphors is for entire transcripts of contextual interviews. Also, it is important to uncover possibly age-universal image-schematic metaphors when only transcripts from younger adults are available. In this case, different parameters are thinkable to guide the selection of found image-schematic metaphors (e.g. number of instances over all transcripts, number of different extractors who found a metaphor, number of participants who used a metaphor). Finally, to apply the method of image-schematic metaphors in the field, an estimation of the inter-rater reliability under realistic conditions is needed. In sum, the study addressed the following two hypotheses:

H1 Younger and older adults do not differ in their technology-*independent* prior knowledge in the form of mental models and show a substantial overlap in their use of image-schematic metaphors in natural language.

H2 The inter-rater reliability of the extraction process of image-schematic metaphors is at least substantial for transcripts of Contextual Interviews.

5.2.1 Method

Participants

Following the typical sample size of Contextual Interviews (Holtzblatt & Beyer, 2017), five younger adults between 22 and 23 years (M = 22.6, SD = 0.55) and five older adults between 54 and 61 years (M = 56.6, SD = 2.70) were recruited via a convenience sample. Participants were compensated with 10 Euros for their participation.

Table 5.5: Demographic description, owned car type and annual driving distance of the participants of study 2.

Participant	Age	Gender	Profession	Car type	Kilometers per year
P1	22	male	student	Nissan Primera	50,000
P2	22	male	student	BMW 730i	12,000
P3	23	male	industry worker	Audi A3	25,000
P4	23	female	student	BMW E30	14,000
P5	23	female	student	VW Fox	3,000
P6	54	male	taxi driver	Hyundai i40	120,000
P7	55	female	nurse director	Ford Focus	20,000
P8	56	female	nurse	Toyota Picnic	3,000
P9	57	male	salesman	BMW 316i Touring	30,000
P10	61	male	truck driver	Mercedes Actros 470	80,000

In each group, two participants were female, and all participants were German native speakers. The five older adults were two nurses, a taxi driver, a truck driver, and a business worker. The younger adults were four students and one industry worker (all still working). Younger adults descriptively showed a lower yearly driving distance of approximately 20,800 kilometers (SD = 18,100km) compared to older adults with 50,600 kilometers (SD = 48,200km). Still, this difference is not significant, $t(5.01) = 1.29$, $p = .25$. One younger adult and two older adults did not own a smartphone. As expected, the other four younger adults used smartphones significantly more minutes per day (M = 217.5, SD = 45.0) than the other three older adults (M = 45.33, SD

= 14.50), $t(5) = 6.25$, $p = .002$, $d = 4.78$. The team conducting the Contextual Interviews and extracting image-schematic metaphors consisted of three male graduate students (not including the author of this work) that were familiar with both methods and had used them previously in multiple projects. They were trained in extracting image-schematic metaphors during their curriculum and again for the project using sample texts and supervision (by the author of this work).

Procedure

The Contextual Interviews took place in the participants' cars with the participant on the drivers' seats and the interviewer next to them. Due to insurance restrictions, the vehicle was not moving. All participants gave their written permission to have their conversation audio-recorded during the Contextual Interview for later analyses. At the beginning of each one-on-one Contextual Interview, each participant answered basic questions on demography and technology use (e.g., smartphones). The main course of the Contextual Interviews followed Holtzblatt and Beyer (2017) and was semi-structured to guarantee the coverage of fundamental concepts (see Appendix A.2.1) while still being open to new topics. For example, as a response for the question "What music do you like", one participant spontaneously began to explain that he often recommended a specific music track to a friend, leading to the next question "Can you show me the concrete steps how do you do that?". Participants were asked to imagine a hypothetical future, more advanced features that were currently not installed in their cockpit. To prevent participants from constraining their ideas, they were instructed not to consider safety issues, for example, due to autonomous driving which would ultimately free them from the driving task. The interviewer directly discussed emerging design ideas with the participant. The Contextual Interviews led to a set of exciting interaction concepts as, for example, throwing music recommendations directly to a friend's preceding vehicle via free-hand gestures. The audio-recorded parts focusing on social communication and entertainment between cars lasted between 30 to 60 minutes (time for introduction, breaks and ending excluded). Only the recordings of these parts were later used during the extraction process of image-schematic metaphors.

Image-schematic metaphors extraction process

In total, 6 hours and 46 minutes of natural language were recorded. On average, Contextual Interviews with younger participants were shorter (34.7 minutes) than with older participants (44.5 minutes). As in study 1, all audio recordings were transcribed on a word-by-word basis. Based on these transcripts, the same team that had conducted the Contextual Interviews also performed the extraction of image-schematic metaphors. They are now referred to as "extractors". Each extractor individually and independently extracted image-schematic metaphors from each of the Contextual Interviews and no abstract concepts were predefined to focus on. To standardise the extraction process, the set of keywords from Löffler, Hess, Maier et al. (2013) was provided and discussed during the training on a sample text (e.g. "an/ab" [on/off] as indicators for CONTACT/NON-CONTACT). To be able to compare the results of the different extractors unbiased by learning effects, the extraction order of the interviews was randomised but the same for all extractors.

In the first step, image schemas of the list of Hurtienne (2011) were extracted from the transcripts (for the procedure, also see Löffler, Hess, Maier et al., 2013). All image schemas were extracted, even for repeated expressions. It was, therefore, possible to extract the same image schema and the same image-schematic metaphor at different locations in the transcripts for identical expressions.

After all extractors had finished,[5] all image-schematic metaphors were aggregated into a single dataset for subsequent analysis. This dataset contained the single extractor, the metaphor itself, the location in the transcript, and the exact word sequence the image-schematic metaphor was based on. Supervised by the author of this work, the three extractors edited and summarised the extracted image-schematic metaphors slightly by removing spelling errors and minimal differences of the extracted image-schematic metaphors (e.g., SYSTEM IS CONTAINER, and SYSTEMS ARE CONTAINERS were merged). The final dataset consisted of 6315 instances of image-schematic metaphor (over the three extractors and ten transcripts). 382 image-schematic metaphors were extracted with at least two instances over all transcripts.[6]

To allow for the comparison between extractors of extracting image-schematic metaphors from one specific location in the transcripts, 2492 word areas were defined manually in the transcripts by the author. Contrary to locating each image-schematic metaphor in all transcripts via only single words, text lines, or paragraphs, this procedure proved to be more flexible in image-schematic metaphors that relied on the context of a keyword. For example, the first extractor annotated SYSTEM INPUT IS IN to the phrase of "schau, hier kann ich meine Zieladresse *ins* System *eingeben*" [look, here I can *in*sert my destination *into* the system] but the second extractor annotated the same image-schematic metaphor to the single word "*ein*geben" [*in*sert]. In this case, one word area "schau, hier kann ich meine Zieladresse ins System eingeben" was generated where both extractors had consistently extracted the image-schematic metaphor SYSTEM INPUT IS IN. Note that by this, multiple image-schematic meta-

[5]Due to the amount of text material that had to be extracted, the extraction process could not be completed in one session. Therefore, each extractor worked independently over a period of three months to finish all transcripts. The self-reported time for extracting metaphors from the transcriptions of the 6.6 hours of audio material ranged from 40 to 70 hours depending on the extractor. Figure 5.2 shows the progress extraction phase. Note that the three extractors differed significantly in their depth of extracting image-schematic metaphors.

[6]This high number of image-schematic metaphors mainly stems from many very specific target domains. For example, different functions of a navigation system led to also different image-schematic metaphors (YOUTUBE IS SURFACE and MUSIC IS SURFACE), which refer to similar, but not identical abstract concepts. Also, some metaphors were specific to single participants. For example, some contextual interviews revealed metaphors that were unique to a participant (e.g. COMMUNICATION BETWEEN TAXI AND HEADQUARTER IS PATH and HEADQUARTER IS CENTER).

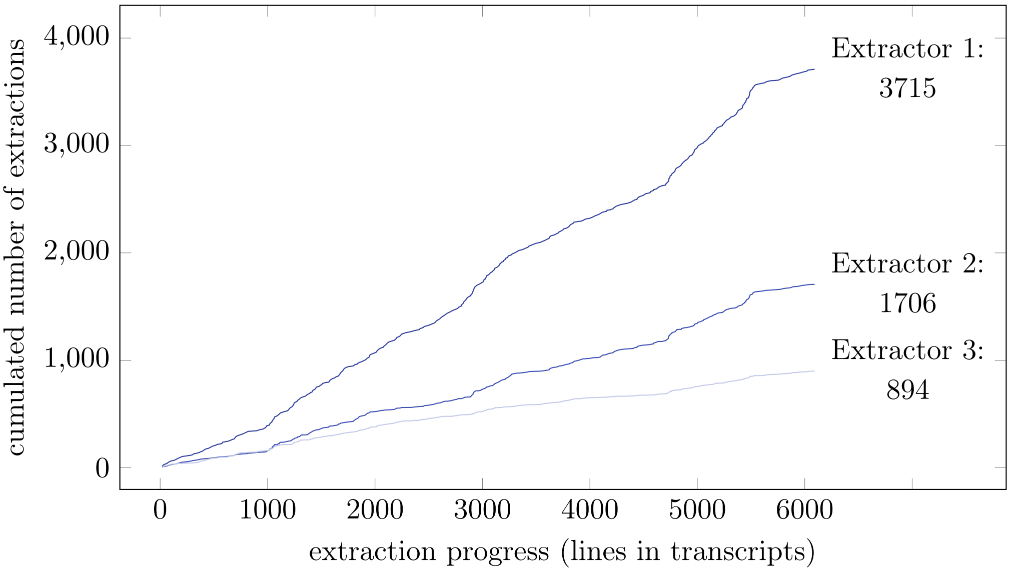


Figure 5.2: Cumulation of extracted instances of image-schematic metaphors per extractor during the extraction process.

phors could be extracted by the same extractor from the same word area. All data analysis was conducted using Excel 2015 and SPSS version 24.

5.2.2 Results from the Extraction Process: Inter-Rater Reliability

The process of extracting image-schematic metaphors from natural language is apt to ambiguity (Hurtienne, 2017a; Löffler, Hess, Maier et al., 2013) and it is essential to provide further insights on how the methodology could be improved in future applications. For this, the following section reports on vital statistics on the image-schematic metaphors during the extraction process.

General frequencies and differences between extractors

In total, the three extractors individually found 6315 instances of image-schematic metaphors to the transcripts of the ten participants. Note that by this procedure, duplicates were possible at the same location by different extractors. As described above, the three extractors worked autonomously and were only supervised during the extraction phase. The three extractors found image-schematic metaphors at 4441 locations in the transcripts. Of all 6315 instances, 58.8% stemmed from extractor 1, more than the 27.0% from extractor 2 and 14.2% from extractor 3 (see also figure 5.2). These differences mainly can be explained by different extraction depth. For example, all three extractors found the image-schematic metaphor AUTOMATIC FUNCTION IS SELF-MOTION in the transcripts (e.g. based on the expression "da *läuft* das dann praktisch durch" [it runs through]). However, extractor 1 found this metaphor at five different locations in the transcripts, but extractors 2 and 3 only at two locations. Another example was the image-schematic metaphor FUNCTION IS SURFACE (e.g. based on the expression "geh mal *auf* Youtube" [go on youtube]), which extractor 1 found at three locations but extractors 2 and 3 only at one location.

Analysing the distribution of image schemas revealed a strong prevalence of the categories CONTAINMENT, FORCE and SPACE. With 12.0%, COMPULSION was the most frequently used image schema, followed by SURFACE, CONTAINER, and UP. Extractors slightly differed in their frequency of extracted image schemas. For example, extractor 3 used CONTAINER in 11.1% of his annotations, more often than extractor 1 (8.9%) and extractor 2 (6.6%). Also, 6.0% of the annotations of extractor 3 fell into the category LOCATION, but only each 1.3% of extractors 1 and 2. This finding emphasises the significance of considering inter-individual differences during the extraction process of image-schematic metaphors.

Inter-rater reliability

A standard measure for the inter-rater reliability between two raters (in this case extractors) is provided by Cohen's Kappa coefficient and can be improved to Fleiss' κ to allow for an estimation of the inter-rater reliability between three or more raters. It represents, therefore, a more robust measure than mere percentages of the agreement since the probability of agreement by chance is taken into account. Fleiss' κ is usually used for a pre-defined set of material that is entirely analysed by all extractors without missing values. However, in this study, the coding process of transcripts was much less structured and – on account of inter-individual differences – not all extractors analysed all words.

Hayes and Krippendorff (2007) discuss different inter-rater reliability measures in the context of text tagging based on five main criteria (measurement of agreement between two or more raters, independence of the number of categories, reliability value on a scale between 0 and 1, applicability for different data levels, computer-based calculation). Their main argument against most reliability measures like percentages or κ indices is that they are not suited for many different categories the raters can choose from. Also, they are difficult to apply in the context of text tagging with a very high number of elements that can be potentially tagged (each word represents an element that could be tagged). To provide a generalised approach for estimating the inter-rater reliability for text-tagging methods, they, therefore, suggest a new measure, called Krippendorff's alpha with the basic formula

$$\alpha = 1 - \frac{D_\mathrm{o}}{D_\mathrm{e}}$$

with D_o as the observed disagreement within the units of analysis (i.e., elements) and D_e as the expected disagreement when the coding is based on chance. The value space of Krippendorff's α is comparable with Cohen's or Fleiss' κ and also ranges between -1 (total disagreement) to 0 (agreement on chance level) and 1 (perfect agreement). Still, due to its foundation on disagreement calculations, it is regarded as more robust, especially when the dataset contains missing data and there is a chance that two of three extractors do not tag a single element (Freelon, 2013). It can be used for a variety of inter-rater reliability problems and embraces several classes of other known reliability coefficients. In the case of three raters, missing data (not all extractors extracted metaphors from all possible locations), and a high number of categories, it is similar to the analysis of variance terms (Freelon, 2013; Hayes & Krippendorff, 2007). In the following sections, both Fleiss κ and Krippendorff's α will be reported to provide a classical as well as a more robust IRR measure. For calculating Krippendorff's α, the tool ReCal3 was used, which is provided by Freelon (2013).[7]

The open-ended content of the Contextual Interviews and the structure of the extraction process must be considered when assessing a realistic inter-rater reliability. In contrast to previous studies where different extractors extracted image schemas from single sentences for given target domains or image-schematic metaphors from pre-selected utterances (Hurtienne, 2011; Löffler, Hess, Maier et al., 2013), the dataset of this study was much less pre-selected and prepared for the extractors. Two judgments had to be made by extractors for each extracted image-schematic metaphors: First, they had to recognise word areas in the transcripts where participants had talked metaphorically about an abstract concept and elect the image schema that had been expressed (e.g., ATTRACTION, COMPULSION). Second, they also had to define the abstract concept (i.e., the target domain, e.g., RECOMMENDATION). Only after the second step, the extractor could annotate an image-schematic metaphor to the transcript. Therefore, it should be taken into account that the inter-rater reliability might be under- or overestimated if these different decisions are not strictly separated. Thus, based on these considerations, two different types of inter-rater reliabilities were calculated for this study.

Two extractors could not match in their extracted image-schematic metaphors at a single location when they had not focused on the same target domain. Thus, only the agreement between extractors with the same target domain at the same location in the transcripts was analysed. At least two of the extractors had to focus on the same target domain at the same location in the transcripts (e.g., RECOMMENDING IS ...). Therefore, the reliability represents the level of disagreement in their source domain (e.g., CONTACT versus PATH). If only two extractors had focused on a target domain at a single location but not the third, the third extractor was treated as a missing value. In 1084 locations, two or all extractors had extracted an image-schematic metaphor with the same target domain. One substantial effect of the target domain filtering was the assimilation of the numbers of annotations between the three extractors. Because extractor 1 showed a high number of target domains at locations that only he annotated but none of the other two, these unique cases were excluded altogether. However, it should be noted that this raised the question of whether design-relevant metaphors could have been excluded by this approach.

Using this dataset revealed a Fleiss' κ of 0.704 for image schemas (observed agree-

[7] Available at (last accessed October 10th 2016): http://dfreelon.org/utils/recalfront/

ment: 0.742, expected agreement: 0.129[8]) and 0.652 for image schema categories (observed agreement: 0.742, expected agreement: 0.258; note that here the expected agreement is higher since the extracting is based on fewer categories than single image schemas and the probability of matching by chance is higher). This value has to be interpreted as a substantial agreement (Landis & Koch, 1977). Krippendorff's α results in comparable values of 0.679 for image schemas (1084 cases, 903 missing data[9], 3252 decisions, observed disagreement: 0.290, expected disagreement: 0.904) and 0.621 for image schema categories (1084 cases, 903 missing data, 3252 decisions, observed disagreement: 0.290, expected disagreement: 0.765). This inter-rater reliability is comparable to other studies that focused on the extraction process of image schemas and image-schematic metaphors from natural language (see Hurtienne, 2011).

Validity of extracted image-schematic metaphors

Image schemas and image-schematic metaphors should be extracted only in case of metaphorical language and explicitly not in case of words describing physical facts and conditions. Still, it is very likely that mistakes occur during the extraction process. Therefore, a third of all instances (2260 out of 6315, 35.8%) was reviewed by an additional expert in image-schematic metaphors assisted by a second reviewer. The first one was the author of this work, and the other one worked for six years with image-schematic metaphors in different projects. For this, two Contextual Interviews per age group were selected. The re-examination was based on the extracted image-schematic metaphors and the corresponding expressions in the transcripts. In case of ambiguity, the decision whether the expression underlying the extraction was metaphorical or not was grounded on the German Duden (Duden, 2017), a dictionary that covers different facets of German words including their origins and different, also metaphorical, meanings.

The analysis revealed that 335 of 2260 reviewed instances of image-schematic metaphors (14.8%) were based on physical instead of metaphorical meanings. For example, the extraction "Smartphone is Surface" was based on word areas as "Ich schaue auf das Handy" [I look at my smartphone]. A smartphone is a real surface, that one can look at physically. Deriving image-schematic metaphors from utterances like this is therefore not justifiable because participants only described the physically obvious and not in a metaphorical way.

Further, 55 of the 2260 instances (2.4%) were duplicates within the same extractor at the same location (e.g., when the same extractor annotated at the same location TOPIC IS COLLECTION and TOPIC IS SURFACE, as well as TOPIC IS COLLECTION AND SURFACE, the combination was an unnecessary duplicate). Also, in 34 of the 2260 instances (1.5%), the extraction was inexplicable to the reviewers. Interestingly, a high

[8]The expected agreement takes into account the probability of agreement by chance, which is very low in this case due to the high number of categories (image schemas), see Hayes and Krippendorff (2007)

[9]Note that Krippendorff's alpha does not exclude missing data listwise but pairwise.

Table 5.6: Frequency of image schema categories per age group.

Category	Younger Adults	Older Adults	Overall
FORCE	27.8%	30.4%	29.5%
CONTAINMENT	28.5%	26.7%	27.3%
SPACE	26.3%	27.4%	27.0%
MULTIPLICITY	12.8%	11.4%	11.9%
ATTRIBUTE	2.4%	2.3%	2.3%
PROCESS	1.8%	1.4%	1.5%
BASIC	0.4%	0.4%	0.4%

number of image-schematic metaphors was irrelevant for the project or interaction design (39.6%). For example, "Meiner Meinung *nach*" [in my opinion] was coded as OPINION IS FRONT-BACK. From a linguistic point of view, it could be worth a debate why we use this FRONT-BACK image schema in this context, but it is useless for the design process per se. Only 943 of the 2260 instances were labelled as metaphorical and relevant extractions (41.7%).

5.2.3 Results on the overlap between younger and older adults

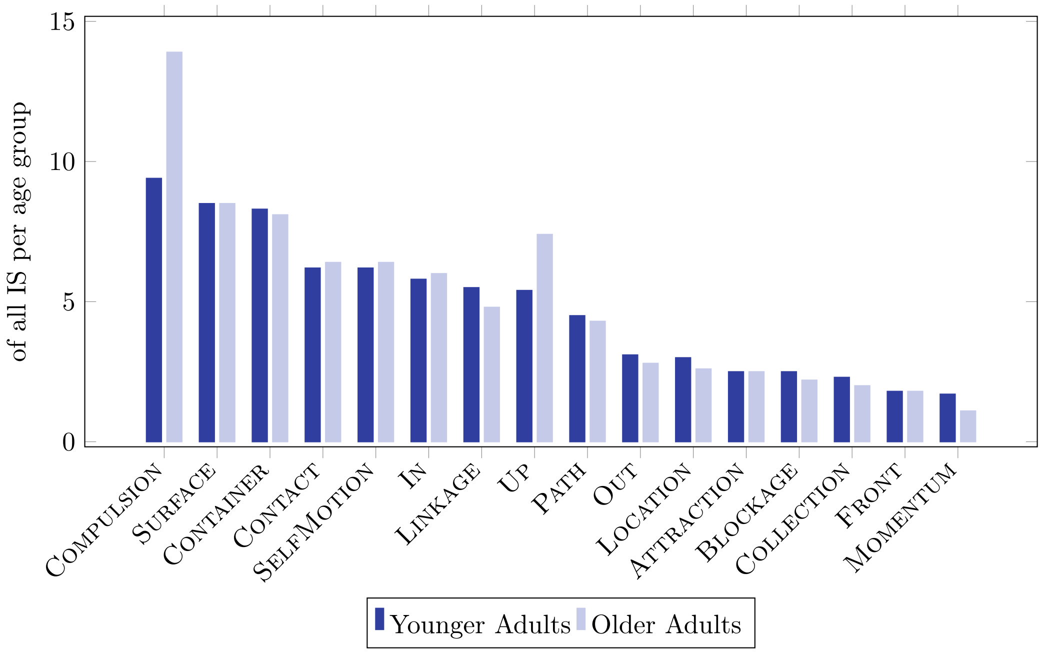
Image schemas per age group

As in study 1, it was expected that younger and older adults would not differ in their frequency of image schemas in their natural language (M. Johnson, 1987). Over the entire corpus of recorded natural language from all descriptions, image schemas from the categories FORCE and SPACE were used most often by both age groups, followed by CONTAINMENT and MULTIPLICITY (see table 5.6). Figure 5.3 shows the most frequent image schemas per age group over all descriptions. Descriptively, younger and older adults showed similar frequencies of the different image schemas, with older adults using slightly more often COMPULSION and UP compared to younger adults. Despite these differences, the two age groups show a similar pattern in the distribution of image schemas in natural language. When considering only image schemas with at least one instantiation over all participants (50 image schemas), both groups highly correlate in their distribution of IS, $r(38) = .971$, $p < .001$. Comparing the ranks of image schemas between both age groups did also not reveal any significant differences, $U = 1713$, $p = .45$.

Image-schematic metaphors per age group

The contextual interviews were not completely standardised and it was desirable for the project that interviews branched into different topics (following Holtzblatt & Beyer, 2017). Due to the semi-structured setting, the Contextual Interviews covered a wide range of topics. In total, the extracted image-schematic metaphors covered

Figure 5.3: Most frequent image schemas used by younger and older adults.



3082 different abstract concepts (and thus target domains). However, only 170 out of 3082 different target domains were a topic in the Contextual Interviews of both age groups (5.5%). 1051 of all target domains only came up in Contextual Interviews with younger adults (34.1%) and 1861 only in Contextual Interviews with older adults (60.4%). However, as became apparent in the later design process, those abstract concepts that were relevant for the project were covered in interviews of both younger and older adults.

To still allow for a comparison between younger and older adults, only abstract concepts were analysed that both groups had repeatedly talked about (i.e., two out of five participants per age group). Thus, if only one younger adult and two older adults talked about a concept, this concept was excluded from analysis. The resulting dataset included 25 target domains with in total 68 image-schematic metaphors (based on 464 instances in the transcripts). This method was different than in study 1 because of two reasons. First, contextual interviews were less standardised than the descriptions in study 1 and not all participants talked about the same topics. The reduction to most relevant abstract concepts was necessary for comparison. Second, less participants were involved in study 2 compared to study 1. Thus, comparing percentage values (five participants per age group) was not justified.

For each of the remaining 25 target domains, the overlap between younger and older adults was calculated by dividing the number of shared image-schematic metaphors by all image-schematic metaphors used for this target domain (similar to study 1). For example, for the concept of CALLING, both age groups only adopted the image-schematic metaphor CALLING IS CONTACT resulting in an overlap-percentage of 100%. Also, both age groups used COMMUNICATION IS LINKAGE, COMMUNICATION IS IN, and COMMUNICATION IS UP, also resulting in an overlap of 100% for the abstract concept of COMMUNICATION. For speech input, the overlap was only 50%, since both age groups used SPEECH INPUT IS IN, but only older adults SPEECH INPUT IS ENABLEMENT. Also, both younger and older adults used the image-schematic metaphor TOPIC IS SURFACE when describing what they were talking about in a conversation. But only older adults used metaphors like TOPIC IS LOCATION, TOPIC IS UP, TOPIC IS CONTAINER, and TOPIC IS IN. Here, the overlap was only 25%. The 25 target domains contained examples with 0 to 100% overlap between both age groups (see also figure 5.4). In total, younger and older adults showed an average overlap of 69.88% (SD = 30.70%) in their use of image-schematic metaphors.

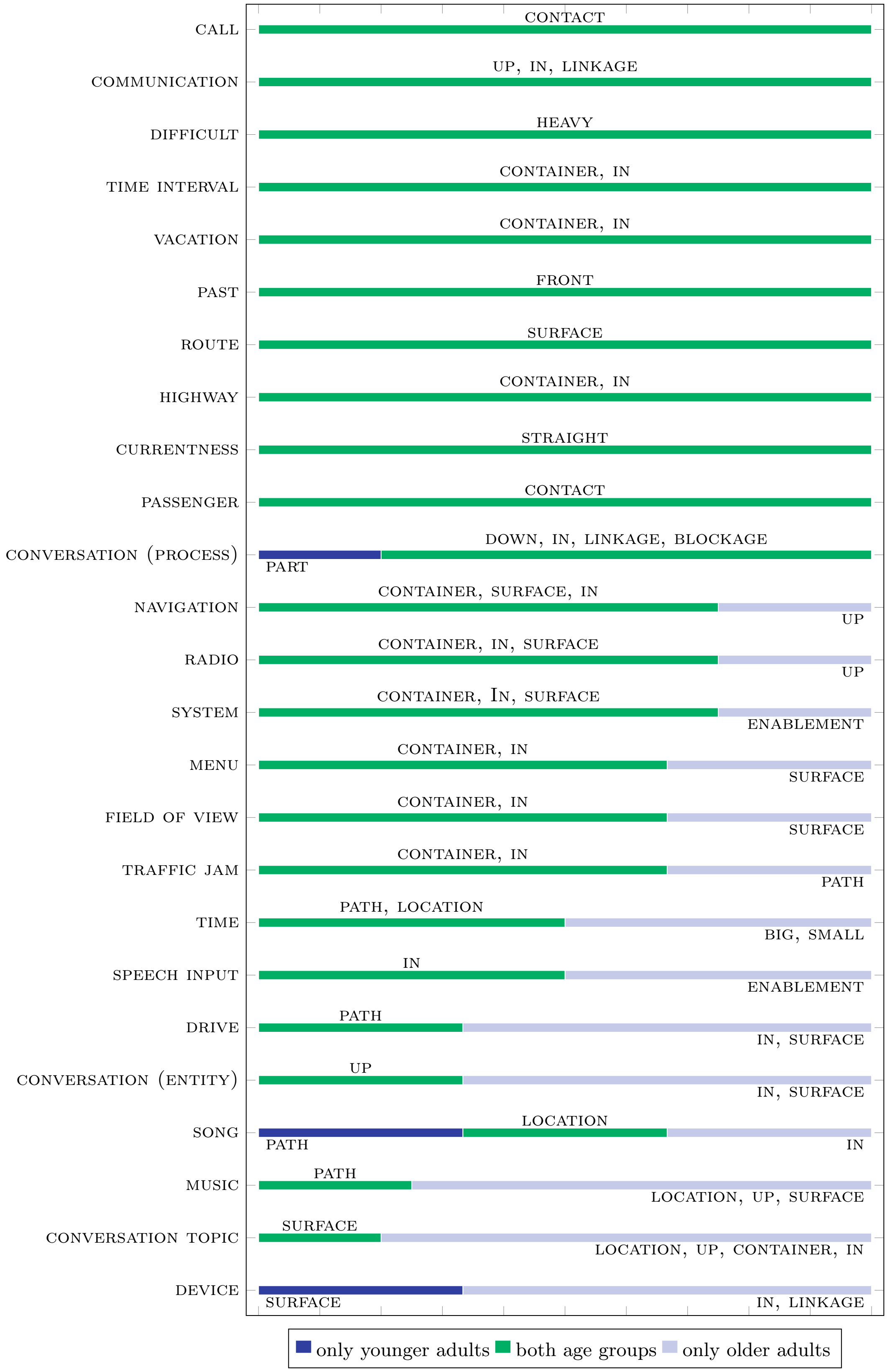
Predicting Older Adults' IS-M from Younger Adults' IS-M

Given the fact that the general overlap between different age groups depends on the target domain of interest, the following section investigates the question which image-schematic metaphors qualify most probable to be age-independent. Of particular interest was how focusing on younger adults as the primary source for image-schematic metaphors would also reveal image-schematic metaphors that would have been extracted from spoken language of older adults, a realistic and vital scenario in user studies using mostly younger adults. For each image-schematic metaphor that occurred in the sample of younger adults, three predictors were extracted. First, the *number of instances* of the image-schematic metaphors over all younger adults' interviews (i.e., the number of extractions per image-schematic metaphor in younger adults). Second, the *number of participants* who used the image-schematic metaphors in the group of younger adults. Third, the *number of different extractors* who had extracted the image-schematic metaphor at least once in all transcripts of younger adults. These parameters served for predicting whether an image-schematic metaphor found in younger adults' transcripts also occurred in at least one older adult's transcript. In total, higher values on each of the three predictors increased the probability that an image-schematic metaphor of younger adults was also found in transcripts of older adults but to different degrees (see also figure 5.5).

Frequency of image-schematic metaphor over all younger adults' interviews. Even if an image-schematic metaphor had been extracted three times in younger transcripts, the probability that the image-schematic metaphor would also occur in older adults' interviews was only 33.0%. Also, out of 1664 image-schematic metaphors in the sample, only 59 image-schematic metaphors were extracted three times in younger adults' interviews. If at least four instances of the metaphor had been extracted from younger adults' interviews, the probability increases up to 57.4%, but only 32 of these cases were found in this study.

Number of younger adults using the image-schematic metaphors. If only one young participant used an image-schematic metaphor, the probability of find-

Figure 5.4: Target domains (y-axis) and their image schemas used by both age groups (above bars) and only younger or older adults (beneath bars).



ing the same image-schematic metaphor also in older adults was only 5.4%. Only 46 image-schematic metaphors were used by at least two younger participants, but 56.5% of them were also used by at least one older adult.

Number of extractors extracting the image-schematic metaphor. Interestingly, the number of extractors who extracted a specific metaphor has almost no predictive value. Even if three extractors extracted the same metaphor (which happened only in 47 out of 1664 cases), the probability of finding the same metaphor also in the data from older adults is only 14.9%.

In this dataset, the highest probability (100%) for finding the same metaphor also in the older adults group was when the image-schematic metaphor had been extracted at least six times or in transcripts of at least three different younger participants. This data suggests that for predicting older adults' use of younger adults' image-schematic metaphors, the number of extractors and the number of instantiations per metaphor are not as important as the number of participants who used the image-schematic metaphors independently. To statistically test this descriptive finding, a binary logistic regression was conducted to analyse the effect of the three predictors on the criterion "image-schematic metaphor shared by younger and older adults". The assumption of absence of multicollinearity between the predictor variables was tested via a correlation matrix among the predictors. The assumption is met if no correlation between predictors is over .90 (Tabachnick, Fidell & others, 2001). The highest correlation was between the number of different participants per image-schematic metaphor and the number of instances per image-schematic metaphor (r = .54). The assumption of absence of collinearity can, therefore, be regarded as fulfilled. Also, the assumption of independent observations and mutually exclusive categories (image-schematic metaphor in both age groups: yes or no) are fulfilled.

Table 5.7 shows the results of the binary logistic regression. The number of different younger participants who used the image-schematic metaphor had the highest impact on the probability that it would also be found in transcripts of older adults. The factor "different participants" is also the only significant predictor. Here, the chance of finding the same metaphor also in the group of older adults increases by the factor 1 to 11.4, if an additional participant in the group of younger adults uses the metaphor. The other two factors "number of instantiations" and "different extractors" are only marginally significant. Taken together, prioritising image-schematic metaphors that were found in transcripts of different participants should be beneficial for age-inclusive interaction design. However, the age-universality of single metaphors (found in the natural language of younger and older adults) should not overrule the relevance for the later design process (functionality). Even when a metaphor is age-universal, but banal or not related to the project, it should not be included in the final set of image-schematic metaphors that is used for interaction design.

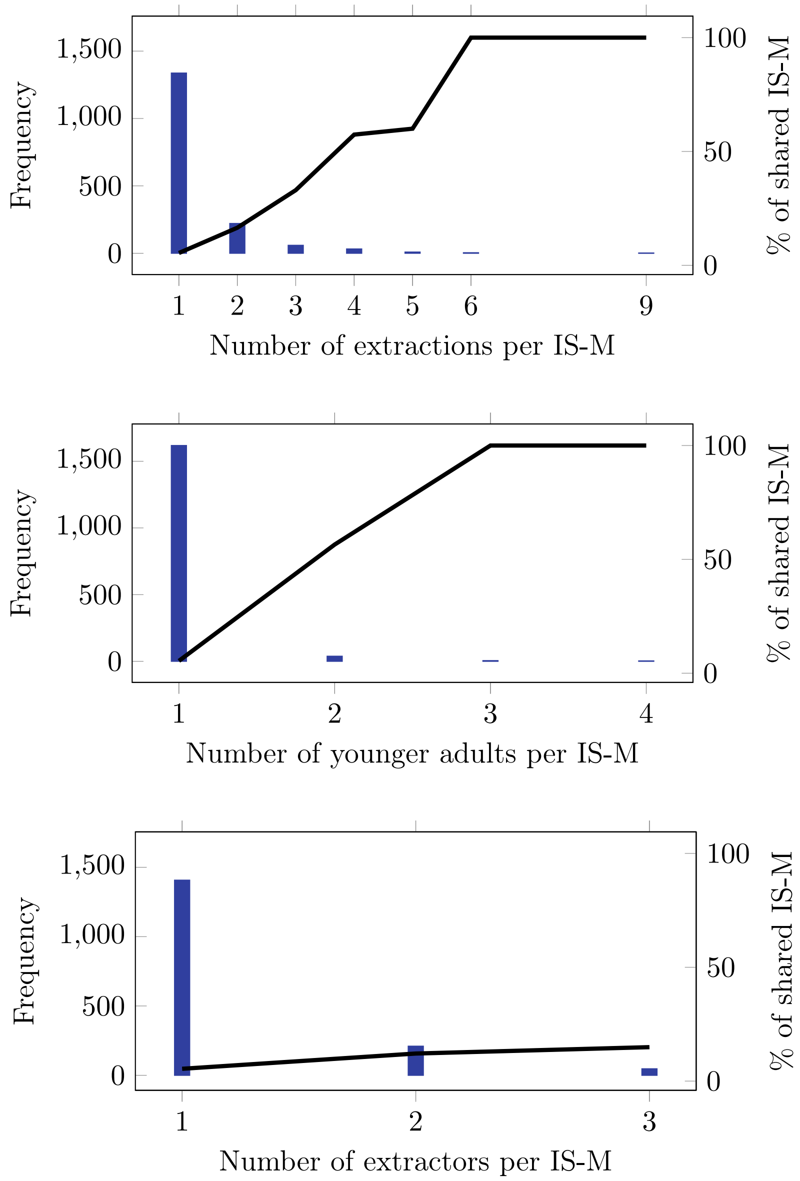


Figure 5.5: Percentage of IS-M used by younger and older adults dependent on the three predictors number of extractions per IS-M (top), number of different younger adults that used the IS-M (middle), and number of extractors per IS-M. Bars show the frequency of IS-M with n extractions (top), IS-M used by n different younger adults (middle), and IS-M extractions by n different extractors (bottom).

Table 5.7: Logistic regression model predicting shared image-schematic metaphors.

Variable	β	SE	Wald	p-value	Exp(β)	95%-CI
Constant	-6.183	1.800	11.798	.001	.002	
Number of instantiations	-.345	.192	3.229	.072	.709	(0.49-1.03)
Different extractors	.904	.466	3.774	.052	2.470	(0.99-6.15)
Different participants	2.432	.855	8.085	.004	11.380	(2.13-60.83)

5.2.4 Discussion

In study 2, a small project team conducted Contextual Interviews with five younger and five older adults and analysed the transcripts for image-schematic metaphors. In a natural and realistic setting of user research, the study was designed to test two hypotheses: First, similar to study 1, a substantial overlap in the occurrence of image-schematic metaphors was expected between both age groups (H1). Second, a substantial inter-rater reliability between the three extractors was expected (H2). Additionally, the study focused on the potential for improvements regarding efficiency and validity of the extracted metaphors. First, the study compared predictors for age-independent image-schematic metaphors. These predictors allow – to a certain extent – to also utilise sets of metaphors that were extracted only from the language of younger adults. Second, due to the subsequent review of the metaphors, the findings reveal some pitfalls that should be addressed in the training of extractors to increase the relevance and validity of the extracted metaphors.

The overlap of image-schematic metaphors between both age groups was substantial and, as in study 1, H1 can be confirmed. However, the overlap was, again, not high enough to recommend only collecting image-schematic metaphors from only one age group for the claim of revealing age-independent mental models. The overlap depends on the abstract concept and can range between 0% and 100%. The final set contained also image-schematic metaphors already reported in literature (e.g., TIME IS ON A PATH, RELATIONSHIP IS LINKAGE), as well as new ones (e.g., INPUT IS ENABLEMENT, CONVERSATION IS CONTAINER). Additionally, both age groups did not differ in their distribution of image schemas in their natural language, as predicted by M. Johnson (1987).

Even though prior technological knowledge was not the main focus of this study, both age groups differed significantly in their exposure to modern technology as shown by their smartphone usage. Since younger and older adults differed significantly in their smartphone use they were expected to differ also in other levels of prior technological knowledge. This is in line with other studies were exposure to single technologies often serves as an indicator for other levels of prior technological knowledge (Blackler et al., 2010).

The three extractors showed a substantial agreement in their extractions when the target domains were held constant. However, because extractors strongly differed in their number of extracted metaphors, H2 was only confirmed partially. In the calculated inter-rater reliability, only image-schematic metaphors were compared where extractors had extracted the same target domain. This approach estimated agreement between extractors if they had to extract image schemas for a predefined target domain at a specific location as in Hurtienne (2011). In general, extractors seemed to disagree mostly between categories (e.g. FORCE, ATTRIBUTE, SPACE) and not within categories.

Also, this study reports an estimation of the validity of the extracted image-schematic metaphors. The extractors were familiar with the method, but not experts. Most extractions were valid image-schematic metaphors, but the majority were not relevant target domains for the project scope. Still, 15% were identified as extractions based on physical and not metaphorical meaning. Thus, false positive metaphor extractions are not the majority of extractions, but they still occur too often to be ignored and should be targeted in the training of new extractors. Also, the method was extremely time-intensive. This fact is not optimal since most of the extracted image-schematic metaphors were removed in later steps because they occurred only once over all transcripts. Even though the focus of this study was not on optimising the extraction efficiency, a critical outcome of study 2 is that extractors must be clear about relevant target domains of the project. Following the results of this study, about half of the resulting image-schematic metaphors will be irrelevant for later project phases if they are not aware of the project priorities and essential abstract concepts of interest. However, this enormous number of irrelevant instances will be reduced in practice, when not all parts of the transcripts must be analysed. Due to the experimental setting of the study and minimum standardisation between extractors, all parts of the transcripts had to be analysed, but focusing on only project-relevant concepts and areas within the transcripts should eliminate also extracting irrelevant concepts and image-schematic metaphors. It is also an open question if extractors should first extract all image schemas and only afterwards extracted image-schematic metaphors. Especially for large amounts of recorded spoken language this approach might be less effective and efficient since extractors have to read each passage twice. However, several limitations of this study have to be taken into account when interpreting and generalising the results.

First, no “oldest older adults” were included in the study. Instead, the age of the sample was rather low, which might increase the overlap between younger and older adults. However, given the context and project scope, this age distribution might be more representative of resource-constrained projects. Taking additionally into account the very different professional backgrounds of both age groups, the assumption was that they also differed in their prior technological knowledge in general.

Second, even if not significantly, there was a tendency for a positive correlation of interview length and age. This phenomenon has been reported earlier (Arbuckle & Gold, 1993). Since the total duration of Contextual Interviews with older adults (222:38 min) was longer than those of younger adults (133:08 min), the total length of transcripts also differed between both groups. As a result, also more image-schematic metaphors were extracted from the group of older adults compared to the group of younger adults. In general, older adults seemed more talkative and wanted to explain in detail how they behaved in the given context and what they would expect from new features. One explanation could be that younger adults chunked their information to a higher degree than older adults. Because the Contextual Interviews conducted in study 2 were a representative example of this method (Holtzblatt & Beyer, 2017), it was decided not to filter the data to assimilate transcription length.

The comparison of difference transcription lengths was plausible in this case because it represents data collected under realistic conditions. Limiting the Contextual Interviews to a preset time would have hampered the ecological validity of the results since the method would have been too constrained and distant from practical application. Still, the fact of different transcription length again render the overlap results as conservative estimations. The variation in the duration of the Contextual Interviews also partially explains why many image-schematic metaphors are specific for older adults and only a few for younger adults.

Third, the extraction process itself is difficult to standardise (Asikhia et al., 2015; Hurtienne, 2011; Löffler, Hess, Maier et al., 2013) and standard measures of inter-rater reliability were not applicable. Not surprisingly, the analysis of the coding process itself revealed big differences between extractors regarding their extraction styles. Even though all extractors focused on extracting image-schematic metaphors from the FORCE, CONTAINMENT and SPACE categories, the amount of extracted image-schematic metaphors varied by a factor of three between extractors 1 and 3. Even though extractors were trained as reported in literature (Löffler, Hess, Maier et al., 2013), they made up their own rules as previously observed (Hurtienne, 2011). An important issue is also the wording of the target domain of the image-schematic metaphor (e.g. communication or calling someone), where different levels of abstraction decreased the agreement between different extractors already in defining the target domain. Contrary to more standardised materials used in other studies (Asikhia et al., 2015; Hurtienne, 2011), where extractors assigned single image schemas to single elements, sentences or observations, the extraction material was larger and more complex in this study. For example, in the studies of Hurtienne (2011), extractors searched for single image schemas in single sentences (e.g. from the FORCE category), the target domain was pre-defined (e.g. a specific task step) and the spoken language material was prepared and split to single fragments to facilitate the extraction process. This setting is not realistic in industry projects and, in study 2, extractors did not have a pre-defined list of user utterances. In sum, standard measures for inter-rater reliability are based on assumptions that were not suitable for the process of extracting image-schematic metaphors from Contextual Inquiries by multiple extractors (e.g., large amount of unstructured material, no predefined set of target domains).

5.3 Summary

Previous studies presumed that instantiations of image-schematic metaphors in spoken language are universal and should be used independently of age to a large extent. However, this assumption has so far remained untested (Hurtienne, 2017a). Showing that both age groups overlap in their use of image-schematic metaphors is a necessity to base age-inclusive interaction design on universal image-schematic metaphors.

In sum, the results of study 1 and 2 show consistently that younger and older adults show a substantial overlap of image-schematic metaphors in their natural language. Both studies reveal similar percentages, even though both follow very different methodologies. The first contribution of these two studies is, therefore, providing for the first time evidence that image-schematic metaphors can reveal age-independent prior knowledge. The majority of investigated target domains (i.e., abstract concepts) are associated with the same source domains (i.e., image schemas) via image-schematic metaphors for both age groups. However, there are also target domains where younger and older adults use utterly different source domains or where the overlap is only partial. The average overlap between younger and older adults between 70% and 75%, therefore, supports the idea of using image-schematic metaphors for age-inclusive design but does not provide a "free ticket" for collecting only image-schematic metaphors from younger adults and claiming that they are also representative for older adults.

If not all image-schematic metaphors are universal per se, a new question arises: what image-schematic metaphors should be the focus during interaction design? Löffler, Hess, Maier et al. (2013) suggest prioritising metaphors with the highest numbers of instantiations. While this may be advantageous for projects focusing on only one age group, the recommendation based on the here reported data is different. From the three predictors "number of instantiations per metaphor", "number of participants", and "number of extractors who extracted a metaphor", only the predictor "number of participants " significantly influences the probability that an image-schematic metaphor found in transcripts of younger adults also occurs in transcripts of older adults. Therefore, especially metaphors that occur frequently, but more importantly over different younger adults should be favoured to maximise the probability of using only age-independent metaphors.

From a methodological perspective, extracting image-schematic metaphors from natural language can be quick but practical or slow but more valid. For example, "metaphors" that in fact described physical circumstances could have been removed by applying the MIPVU procedure of Steen et al. (2010) (see section 4.4.1). For linguistic research, this slow but thorough approach would be needed. Since the contribution of this work focuses on the field of HCI and application, this procedure would be even more time-consuming than the extraction process applied in this thesis. In HCI projects, efficient approaches are needed rendering the extraction procedure suggested by Hurtienne, Klöckner et al. (2015) as most feasible in practice. In study 2, the extraction process took – depending on the extractor – between 40 and 70 hours for 6 hours and 46 minutes of audio-recorded contextual interviews. In other words, extracting image-schematic metaphors from an hour of contextual interviews will require between 6 and 11 hours when no abstract concepts are pre-defined for the extractors. Study 1 did not track the time that was needed for the extraction process.

Study 2 revealed 382 metaphors of which were only approximately 42% about relevant abstract concepts in the project (162 relevant metaphors with more than two instances extracted from transcripts of 396 minutes). Study 1, on the other hand, revealed 137 metaphors of which were *all* about relevant abstract concepts (only relevant concepts were described in the transcripts; 137 relevant metaphors with more than two instances extracted from transcripts of 430 minutes).[10] In sum, the method of describing abstract concepts focuses the extraction process on only extracting image schemas for relevant concepts, relieving the extractors from formulating their own target domains of the metaphors. Comparing the two methods on this basis, using descriptions of abstract concepts might be favoured over complete transcripts of contextual interviews.

The overall conclusion of this chapter is that we can expect younger and older adults to possess a common technology-independent prior knowledge and some shared mental models in the form of image-schematic metaphors. Still, to increase both the probability of finding universal image-schematic metaphors and the efficiency of the extraction process, the following methodological recommendations should be considered in practice:

Age-diverse language sample Consider extracting image-schematic metaphors from the language of both younger and older adults. Image-schematic metaphors extracted from the language of younger adults have a high chance of overlapping with that from older adults. However, younger and older adults do not overlap perfectly, and they represent some abstract concepts differently.

Prioritise Focus on image-schematic metaphors that are frequently not only in one participant but over a variety of different participants when only the natural language of younger adults is available.

Training of extractors: project relevant target domains Define and communicate project relevant target domains or abstract concepts. For example, define a number of core use cases (found during the Contextual Design process) and only look at utterances that refer to these. Otherwise, many irrelevant image-schematic metaphors will be extracted from the user interviews, wasting approximately half of the enormous time resources.

Cross-validation between extractors Extractors apply different styles, extracting too many or too few image-schematic metaphors in the worst case. Repeated cross-validation of single extracted passages between the extractors and their extraction depth could increase consistency across extractors.

Training of extractors: physical vs metaphorical Explicitly separate physical and not physical meanings of words. Especially for large amounts of transcripts,

[10]Comparing the two methods in terms of relevant image-schematic metaphors per minute reveals similar outcomes (transcribed interview time needed for one relevant image-schematic metaphor). Study 1 (descriptions) 3.1 minutes for each metaphor. Study 2 (contextual interviews): 2.5 minutes for each metaphor.

extractors will tend to simplify their process and stick to signal words. However, if these signal words are not meant metaphorically in each context, this will bias the results.

Chapter 6

Interaction Design with Image-Schematic Metaphors

Chapter 5 showed that younger and older adults share a substantial percentage of image-schematic metaphors. This form of prior knowledge is largely technology-independent and thus could lead to age-inclusiveness when incorporated into interaction design.

Additionally, image-schematic metaphors could serve as a source of inspiration during the design process (see chapter 4), stimulate the designers' creativity and lead to innovative interfaces. Image-schematic metaphors in the design process, thus, might lead to both innovative and age-inclusive HCI (Hurtienne, Klöckner et al., 2015). If interaction designers recognised the applicability and potential of the new method, the new method could contribute a powerful tool for future interaction design. However, previous work (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013) does not provide evidence that the new method is perceived as equally applicable but more creativity stimulating compared to other methods. Especially methodological weaknesses like missing baselines challenge the significance of positive feedback on the method reported in the literature.

The following chapter focuses on the interaction designer's perspective while addressing weaknesses of previous work. Two studies were conducted to test whether image-schematic metaphors are as applicable as an industry standard (a baseline) in interaction design and whether the potential for creativity stimulation holds true for interaction designers.

6.1 Integrating Image-Schematic Metaphors in the Design Process

Image-schematic metaphors can guide interaction design in different realms such as graphical, tangible and hardware-based user interfaces (Asikhia & Setchi, 2016;

Hurtienne, 2011; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013; Winkler et al., 2016), but also in mapping abstract concepts to gestures or full-body interactions (Antle et al., 2009; Hurtienne et al., 2010). Notwithstanding these many applications, literature is short on the concrete guidance of integration of image-schematic metaphors in the process of interaction design. As one exception, a set of instruments that support the extraction and application of image-schematic metaphors in industry projects is available by Löffler, Hess, Maier et al. (2013). But in sum, reported recommendations often originate from case-studies.(Antle et al., 2009; Hurtienne, Löffler et al., 2015; Löffler, Hess, Maier et al., 2013). For example, Hurtienne, Klöckner et al. (2015) describe the full process of implementing image-schematic metaphors into a User-Centered Design process (Holtzblatt et al., 2005). They evaluated both the design process (from the perspective of the team) as well as the outcome of this process, namely the resulting graphical user interface. In this study, the team gave positive feedback on the method of image-schematic metaphors. Also, participants of the prototype-evaluation rated the final prototype as intuitive and innovative. Most importantly the results of the evaluation were not correlated with age. The authors saw the final prototype as innovative, intuitive to use, and age-inclusive and attributed these findings to the inclusion of image-schematic metaphors in the design process. In sum, this finding is promising for the method of image-schematic metaphors.

However, previous work like this has focused mainly on qualitative feedback from single project teams, included only small sample sizes, and rarely compared the design process to a baseline (Löffler, Hess, Maier et al., 2013; Löffler et al., 2014), even though comparisons are essential for interpreting the results and feedback. More often than not, the design process with image-schematic metaphors has been carried out and evaluated without comparing it systematically to other approaches.

This chapter tries to address this shortcoming and reports on two studies where teams created prototypes with or without image-schematic metaphors. Importantly, the baseline conditions in both studies allowed for interpreting the feedback, and controls for novelty effects of this method. Study 3 followed a standardised experimental design with a larger sample and three design conditions. Study 4 transferred image-schematic metaphors into a design agency with professional interaction designers. In sum, the two studies of this chapter contribute detailed and quantifiable information for specific characteristics of the design process. These results provide the basis for deriving concrete recommendations for improving and standardising the methodology of image-schematic metaphors in interaction design.

6.2 Study 3: Wall Walk with Image-Schematic Metaphors

As described in chapter 4, image-schematic metaphors should contribute a source of inspiration for designers and foster creativity. The main advantage of image-schematic metaphors is that designing with this form of metaphors provides a source for inspiration but at the same time keeps solutions in line with the universal mental model of users independently of their age. Theoretically, designing with image-schematic metaphors should lead to innovative and age-inclusive interfaces (Hurtienne, 2017a), an assumption that will be tested in chapter 7.

Study 3 compared a standard design method in the form of an affinity diagram (see chapter 4; Holtzblatt & Beyer, 2017; Hurtienne, Klöckner et al., 2015) to both the method of image-schematic metaphors as well as the combination of affinity diagram and image-schematic metaphors. In all three conditions, participants conducted a Wall Walk, which is a standard procedure for brainstorming design ideas as a basis for innovative design solutions (Holtzblatt & Beyer, 2017).

The underlying motivation for study 3 was twofold: first, it focused on the perspective of the designer. The design process itself is a principal research object to draw full potential of image-schematic metaphors in interaction design. Previous work showed first positive feedback from design teams (Hurtienne, Klöckner et al., 2015; Hurtienne, Löffler et al., 2015; Löffler, Hess, Maier et al., 2013). However, the literature lacks data from an experimental evaluation of designing with image-schematic metaphors and a baseline like standard affinity diagrams in Contextual Design (for a case-study on the implementation of image-schematic metaphors in Contextual Design see Hurtienne, Klöckner et al., 2015). Second, study 3 generated prototypes of user interfaces that were either developed with or without image-schematic metaphors. As described later in chapter 7, a visual designer further developed the interfaces of study 3. These interfaces were evaluated in a user study regarding perceived innovation and age-inclusiveness (see study 5 in section 7.1).

Zimmerman et al. (2007) state that "there can be no expectation that two designers given the same problem, or even the same problem framing, will produce identical or even similar artifacts". As a consequence, experimental research in the field of design research is challenging. For example, results are far more prone to random effects than in classical behaviour sciences (Dinar et al., 2016). Even though studies in the field of design research sometimes include only small sample sizes (e.g. only a dozen participants, Chulvi et al., 2012), sample sizes of 10 participants per experimental condition can be regarded as already large (e.g. Linsey et al., 2011; L. C. Schmidt et al., 2010; Worinkeng et al., 2013). Additionally, dependent variables for measuring the design outcome vary between studies, especially when the method promises to support innovative design solutions (for an overview of experimental approaches in design research, see Dinar et al., 2016). Often reported dependent variables in design research are characteristics of the *design solutions* (e.g. quantity, quality, novelty and variety of design solutions) and of the *design process* itself (e.g. quality of communication, participants' engagement; Dinar et al., 2016; Silva & Read, 2010).

The collected data covered a wide range of subjective as well as objective measurements. Since the resulting user interfaces were evaluated in a user study (see chapter 7), the focus in study 3 was on the designer's perspective and the designer's perception of various facets of the design process. To be applicable and accepted by designers, the method of image-schematic metaphors should be as comprehensible, easy to apply, and helpful as a standard method like an affinity diagram. Also, it should not affect the perceived quality of the teamwork (as one of the core pre-requisites for stimulating innovative design solutions in Contextual Design Holtzblatt & Beyer, 2017). Additionally, it was of interest how participants would rate the potential of image-schematic metaphors to stimulate creativity as well as their outcome of the design process: the user interfaces should be perceived as more innovative when the design process was based on image-schematic metaphors.

Introducing image-schematic metaphors to the design process might change other aspects as well. On an exploratory basis, the engagement of participants was also measured in the form of the participants' experience of flow (Csikszentmihalyi, 1997) and positive and negative emotions (Watson, Clark & Tellegen, 1988). Finally, the measured variables also included participants' perception of their mental workload.

In essence, study 3 stated two hypotheses on changes of the design process *itself*:

H1 The method of image-schematic metaphors is rated as applicable as an industry standard design method like an affinity diagram.

H2 Designing with image-schematic metaphors does not change key characteristics of the design process itself compared to a standard design method like an affinity diagram: the experience of flow, teamwork, positive and negative emotions, mental workload.

Also, study 3 stated two hypotheses on the *outcome* of the design process:

H3 During a Wall Walk, image-schematic metaphors (alone or in combination with a standard design process like affinity diagram) increase the number of design ideas compared to an affinity diagram alone.

H4 Participants perceive the design process with image-schematic metaphors as more creativity stimulating compared to a standard design process like an affinity diagram alone.

6.2.1 Pre-study

The investigation of the design process in a standardised manner is challenging. A small pre-study was conducted to ensure that the experimental design (instructions, procedures and collected data) would cover the most relevant facets of the design process while at the same time changing the process to a minimal extent. Insights from the pre-study served as the basis for the methodology of the main study reported in this section. In the pre-study, 14 participants (students of human-computer interaction and media communication) individually created paper prototypes (tablet-size) for three use cases in the domain of online-banking. For each use case, they followed a different design method (each printed on sheets of paper): a) affinity diagram alone (see Holtzblatt & Beyer, 2017), b) extreme characters (see Djajadiningrat, Gaver & Fres, 2000), or c) image-schematic metaphors (see Hurtienne, Klöckner et al., 2015). All participants completed all three conditions (within-subject design; 40 minutes design time per condition excluding instructions and questionnaires).

The image-schematic metaphors stemmed from language analysis of contextual interviews with three younger and three older adults (see also the pre-study of study 1 in section 5.1.1). The insights of the procedure of this pre-study were distilled into three primary requirements for the main experiment. First, the method of using an affinity diagram benefits most from a Wall Walk. A too small affinity diagram – feasible in a standardised experimental setting – with too little time for a Wall Walk does not draw its full potential. Instead, the experimental procedure must grant the Wall Walk enough time and the affinity diagram enough space to allow for a proper baseline. Second, individually created prototypes are prone to individual influences, and the communication about the intended interaction design as well as the degree of structure cannot be measured without grouping participants into small design teams. Third, an experimental within-subjects design (all participants complete all three design conditions) is efficient but massively limits the external validity of the results for two reasons. On the one hand, participants can accumulate sources of inspiration between conditions (carry-over effects). On the other hand, the time per condition is limited when each participant must complete multiple conditions. Especially in the case of creating prototypes, a between-subjects design should be favoured, and participants should have enough time to immerse themselves in one design task. In sum, the pre-study also allowed for several improvements of the data collection (questionnaires, instruments) and instructions.

6.2.2 Method

Sample In the main study, 36 participants were recruited. 24 of them were enrolled in a degree course of human-computer interaction and 12 in information and communication design. Their age ranged from 18 to 31 years (M = 23.4, SD = 3.2) and 22 were male (61%). On average, participants had previously worked in 2.1 projects (SD = 1.9) where they had been actively involved in interaction design (e.g., paper prototyping, affinity diagramming). Participants either received 35 Euro or partial course credit for their participation. Because no effect size could be estimated a priori, the sample size followed other studies in the field of design research with often even fewer participants (Chulvi et al., 2012; Dinar et al., 2016; Zimmerman et al., 2007).

Experimental Design The study realised a between-subjects design with the main independent variable "design method". On account of the duration of the experiment and expected carry-over effects (see pre-study 6.2.1), each participant thus completed only one method. The two methods of affinity diagram and image-schematic metaphors were either used alone or combined, resulting in the following three levels of the factor "method":

- affinity diagram only
- affinity diagram & image-schematic metaphors
- image-schematic metaphors only

Two different use cases were selected that represent common functionalities in online-banking to standardise the design task and to extend the range of prototypes: a) "Completing a transaction or a standing order" and b) "Administration of different bank accounts". The use cases stemmed from the pre-study (the third use case of the pre-study was merged into the other two use cases; see section 6.2.1, for the instruction, see appendix A.3.1). Participants were distributed equally on the two use-cases.

Material - design methods The affinity diagram resulted from the six contextual interviews with younger and older adults from the pre-study (see section 6.2.1). The size of the affinity diagram was limited to ensure the feasibility of a small Wall Walk during the experiment (see figure 6.1). The affinity diagram consisted of three main categories (e.g. bank account administration), 5 subcategories (e.g. bank transactions), 18 low-level insights (e.g. "For a bank transaction, I order my bank to transfer money from my bank account to another bank account"), and 94 concrete user utterances and observations (e.g. "Und das braucht ja dann ein paar Minuten bis es auf deinem Konto elektronisch gelöscht ist und auf deinem anderen Konto elektronisch wieder draufgesetzt wird."). In total, the affinity diagram consisted of 120 notes, which is smaller than in most projects (typically 250 to 500 Holtzblatt & Beyer, 2017). However, this size allowed for a small Wall Walk and at the same time constrained the time participants needed to immerse themselves in the data. Eleven image-schematic metaphors were used that were identical with the pre-study (see table 6.1). Without user utterances, they were placed on the wall to allow for attaching design ideas to them. In all three conditions, two personas (one young and one old) were placed on the wall to illustrate an age-diverse target user group (see figure 6.1). Integrating personas in the design process can be regarded as a standard method in interaction design (Dark Horse Innovation, 2016). Personas can avoid interaction designers to design solely for themselves and their peer group (Holtzblatt & Beyer, 2017).

Material - collected data Several variables were measured before, during and after

Table 6.1: Image-schematic metaphors used in study 3. Presented in German to the participants.

TRANSACTION IS COMPULSION ON PATH
INCREASING BANK BALANCE IS INCREASING SUBSTANCE IN CONTAINER
MONEY IS SUBSTANCE
WITHDRAW MONEY IS DOWN
SELECTING RECIPIENT IS CONTACT
BANK IS CONTAINER
BANK ACCOUNT IS CONTAINER
BANK ACCOUNT WITHOUT MONEY IS EMPTY
INTEREST AND MONEY IS LINKAGE
MUCH MONEY IS BIG
STANDING ORDER IS CYCLE

the design process. Demographic data and data on the participants' previous experiences with the single design methods were collected via a pre-questionnaire (see appendix A.3.3). This questionnaire also included information on prior knowledge with online banking in general.

On an exploratory basis, during the experiment, the subjective mental effort questionnaire (SMEQ, Zijlstra, 1993) was applied which is a short but well-performing indicator for mental workload (Sauro & Dumas, 2009). Also, the NASA-TLX was used to explore possible influences of other factors such as time pressure or fatigue during the experiment (Hart, 2006; Hart & Staveland, 1988). The PANAS (Positive Affect and Negative Affect Scales) served for collecting data on the participants' emotional experience (positive and negative) during the experiment (Watson et al., 1988).

After the experiment, the design process itself was rated using 7-point Likert-scales on the following dimensions (see appendix A.3.4): teamwork, general performance, structure. Additionally, participants rated the design method regarding its potential to stimulate creativity and innovative solutions and how helpful it was in general. Also, participants rated whether they created many different, spontaneous design ideas or a few, elaborated ones. Also, participants rated their flow experience during the design process based on the questionnaire of Moneta (2012) that covers the four main facets of flow: challenges of the activity, one's skills matching the activity, the importance of the activity, and satisfaction about doing the activity. Finally, participants could also write down qualitative feedback. In conditions with image-schematic metaphors, a second questionnaire was given to the participants to collect data on aspects of the design process that were specific to image-schematic metaphors (see appendix A.3.8). Here, participants gave qualitative feedback on the advantages and disadvantages specific to the method of image-schematic metaphors.

Procedure Participants were grouped in teams with two members. It was not con-

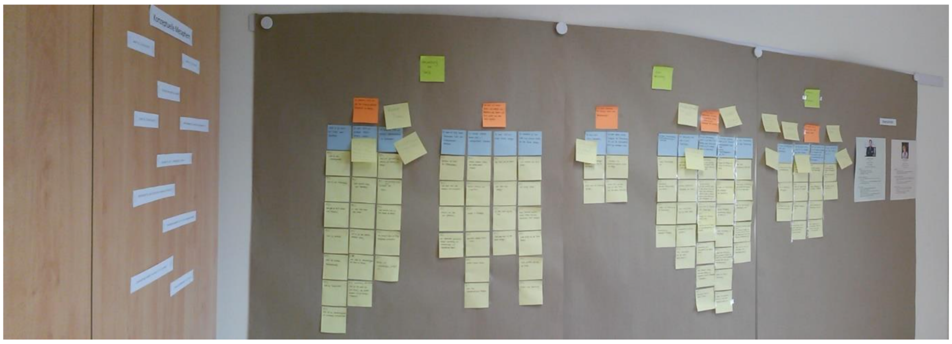


Figure 6.1: Setting of study 3. Left: Image-schematic metaphors. Middle: Affinity Diagram. Right: Personas.

trolled whether they had worked together on previous projects, but the probability for this was expected to be very low due to the large pool of participants they had been recruited from. After arriving at the lab, the two participants signed informed consent for participation. Here, all participants also agreed on the collection of video and audio data of the entire experiment. A video-camera was placed in one corner of the room to record each session. Participants also completed a set of questionnaires on demographics, prior experiences with interaction design, affinity diagramming, and image-schematic metaphors (see section "Material").

Each team was assigned one method and one use case. The experimenter introduced participants to the study and the method of their experimental condition. The duration of the introduction was kept constant between conditions. In the condition of "affinity diagram only", the User-Centred Design process including the structure of the affinity diagram was explained. In the conditions with image-schematic metaphors ("affinity diagram & image-schematic metaphors" and "image-schematic metaphors only"), the User-Centred Design process, the origin of the image-schematic metaphors and basic examples for their instantiations were explained. The instruction phase lasted approximately 20 to 25 minutes including questions. As soon as participants had no further questions on the main procedure of the experiment, participants completed a SMEQ-scale and a PANAS-questionnaire (see section "Material"). During the instruction phase, also the use case was explained. Note that this was the last time, the experimenter answered fundamental questions on the methodology to avoid influencing or biasing the design process in neither direction. Regardless of the condition, the procedure after the instruction comprised three parts and times were held constant. The three phases are summarised in Table 6.2.

In phase 1, participants completed a Wall Walk following the instructions of Holtzblatt and Beyer (2017). The primary purpose of a Wall Walk is to brainstorm design ideas focusing on possible interaction designs for the given use case. The team was instructed to generate as much design ideas as possible for the use case. In all conditions, participants wrote or drew each design idea on a single Post-It™, and attached it to the affinity diagram or the prints of the image-schematic metaphors. Participants

Table 6.2: The three phases of study 3.

Phase	Description	Duration
1	Conducting wall walk, short discussion	35 min
2	Prototyping (alone)	50 min
3	Prototyping (team)	55 min

also summarised their most promising design ideas to hot ideas (Holtzblatt & Beyer, 2017) before entering the next phase. After a standardised duration of 35 minutes, each participant completed a SMEQ-scale and PANAS-questionnaire. Except for the provided "Wall" (affinity diagram, image-schematic metaphors or both), conditions did not differ in their procedure.

In phase 2, the task for each participant was to create one paper prototype for the given use case. For this, participants could use a variety of design materials ranging from the iPad-templates and a set of pens to Post-Its, scissors, foil and other typical prototyping materials. The creation of the first prototype stopped after 50 minutes and participants completed a SMEQ-scale and PANAS-questionnaire.

In phase 3, participants were asked to, together, integrate their two prototypes into a final one. Participants were instructed to bear in mind that the solution should be a) usable by younger and older adults and b) innovative in the end. The reason for creating one final combined prototype was that participants were likely to vary in their design skills. Compared to only one designer at a time, teams of two were expected to be able to compensate for weaknesses of single team members. Additionally, the activity of a Wall Walk is usually a team activity and benefits from interaction between team members. Participants worked 55 minutes as a team and had to integrate the best design ideas from both individual prototypes into one final prototype. Afterwards, they completed a final SMEQ-scale and PANAS-questionnaire.

Finally, participants completed a last questionnaire on their experience of the design process. This questionnaire included questions on the perceived quality of the teamwork, the stimulation of creativity and the structure provided by the method (the one, they had worked with), but also about the quality of the affinity diagram. Furthermore, participants completed a NASA-TLX-questionnaire (see also section "Materials"). For conditions including image-schematic metaphors, questions specific to image-schematic metaphors were asked. Also, they rated the quality of their prototypes regarding how innovative, intuitive to use and age-inclusive they perceived it. Finally, they provided qualitative feedback on the design process in a short post-interview. The study lasted between 3.5 and 4 hours per team including instruction, the three design phases and data collection before, during and after the experiment.

6.2.3 Results

Data analysis was conducted using SPSS version 24 and, if not stated otherwise, alpha-level was set to .05. Degrees of freedom might differ between variables due to correction (e.g., because of violation of assumptions like variance homogeneity). Results will be reported per participant and not per design team. Comparison between the three method conditions will be reported over both use cases.

The self-reported experience in interaction design did not differ between the three conditions, $F(2, 33) = 1.42$, $p = .26$. In general, the own experience was rated as lower ($M = 3.1$, $SD = 1.6$) than the scale mean (4), $t(35) = 3.4$, $p = .002$. Also, the quality of the Affinity Diagram was rated on average with 5.5 ($SD = 0.9$) which is significantly higher than the scale mean (4), $t(23) = 8.3$, $p < .001$. The self-reported experience with online-banking did also not differ significantly between conditions, $F(2, 33) = 1.45$, $p = .25$, and was not rated significantly higher or lower ($M = 4.0$, $SD = 2.3$) than the scale mean (4), $t(35) = .07$, $p = .94$. In conditions including image-schematic metaphors, the participants' self-reported prior experiences with image-schematic metaphors in the design process did not differ, $t(22) = 0.46$, $p = .65$, but was in general rated significantly lower ($M = 2.2$, $SD = 1.8$) than the scale mean (4), $t(23) = 5.1$, $p < .001$.

Table 6.3: Quantitative ratings of the design process. Mean and [SD]. AD: affinity diagram. IS-M: image-schematic metaphors. Poles of all questions (except on "Flow"): Not: 1 to Very good: 7.

Dimension	Question	AD	AD & IS-M	IS-M
Comprehensiveness	How easy was the method to apply?	5.3 [1.2]	5.7 [0.9]	5.6 [1.1]
Structure	How structured was the design process?	5.1 [1.4]	4.8 [1.2]	5.7 [1.1]
Support	How did the method support the design process?	6.0 [0.7]	5.3 [1.0]	5.2 [1.2]
Creativity	How well did this method inspire your creativity?	5.7 [1.4]	5.0 [1.0]	5.6 [1.1]
Teamwork	How well functioned the teamwork?	6.5 [0.8]	5.5 [1.2]	6.4 [1.2]
Flow (Low: 0 to High 9)	How challenging was the activity?	5.3 [1.8]	4.9 [1.9]	5.8 [1.9]
	Your capabilities for this activity?	5.1 [2.5]	6.4 [1.2]	5.4 [1.9]
	Was the activity important for you?	6.4 [1.8]	4.6 [2.4]	5.0 [2.4]
	Were you satisfied with your achievement?	6.5 [2.1]	5.6 [2.3]	6.8 [1.9]

Data on participants' perception of the design process itself. Table 6.3 reports the results of the quantitative questionnaires. No significant differences were found for the ratings of the comprehensibility, structure, support, and creativity between the three conditions. However, the reported teamwork differed significantly between conditions, $F(2, 33) = 3.30$, $p = .049$ (nonparametric: $H(2) = 7.8$, $p = .02$), with the condition "affinity diagram & image-schematic metaphors" rated lower compared to the other two conditions.

Descriptively, participants rated "affinity diagram and image-schematic metaphors" worse than the other two conditions in the dimensions of creativity stimulation and teamwork. Also, participants rated the design process as least structured in this con-

Table 6.4: Quantitative ratings for the two conditions including image-schematic metaphors ("affinity diagram & image-schematic metaphors" and "image-schematic metaphors only"). Mean and [SD], scales from Not: 1 to Very: 7. AD: affinity diagram. IS-M: image-schematic metaphors

Question	AD & IS-M	IS-M
How well did you know the method of IS-M prior to the experiment?	2.0 [1.4]	2.3 [2.1]
How well did the method of IS-M structure the design process?	4.8 [2.1]	5.3 [1.1]
How well did the method of IS-M inspire your creativity?	5.3 [1.3]	5.3 [1.4]
How helpful was the method of IS-M?	5.2 [1.3]	5.3 [1.4]
How easy to understand was the method of IS-M?	5.2 [1.0]	5.4 [1.6]

dition. Still, "affinity diagram & image-schematic metaphors" was rated as the most applicable method. Regarding the four flow-related questions, "affinity diagram & image-schematic metaphors" was rated descriptively higher for perceived own capabilities but lower on the dimensions of challenge, importance, and satisfaction. This is in line with the other questions and shows that the workflow in the combination of "affinity diagram & image-schematic metaphors" was not as smooth as in the two "pure" conditions.

The cognitive load, operationalised via the SMEQ, increased with each of the four phases during the experiment, $F(2.7, 88.7) = 14.28$, $p < .001$, $\eta_p^2 = .30$ (within-subjects, degrees of freedom corrected due to the violation of the sphericity assumption). However, the condition itself did not affect the perceived cognitive load, $F(2, 33) < 1$. Analysis of the PANAS-questionnaire and the NASA-TLX did not reveal any significant differences between conditions or phases (also not on the subscales).

In the conditions "affinity diagram & image-schematic metaphors" and "image-schematic metaphors only", participants completed a number of questions specific to image-schematic metaphors (see table 6.4). Descriptively, the method of "image-schematic metaphors only" was rated as slightly more structuring the design process, but no significant differences between image-schematic metaphors alone or in combination with the affinity diagram were found in the quantitative measures.

Data on the outcome of the design process. The number of design ideas per design team significantly differed between design conditions, $H(2) = 12.02$, $p = .002$. Design teams in the condition "image-schematic metaphors only" had most design ideas attached to the wall ($M = 70.1$, $SD = 33.0$), more than in conditions "affinity diagram only" ($M = 21.8$, $SD = 2.7$) and "affinity diagram and image-schematic metaphors" ($M = 18.7$, $SD = 3.3$). However, the number of final hot ideas that were collected and discussed after the Wall Walk was similar between "affinity diagram only" ($M = 6.5$, $SD = 2.2$), "affinity diagram and image-schematic metaphors" ($M = 5.5$, $SD = 2.0$) and "image-schematic metaphors only" ($M = 5.0$, $SD = 1.7$). Thus, only the number of design ideas but not the summarised hot ideas differed between conditions.

Participants descriptively rated the final outcome of the design process – that is, their own prototypes – in the condition “image-schematic metaphors only” as more innovative (M = 4.83, SD = 1.75) than in the conditions “affinity diagram only” (M = 4.25, SD = 1.54) and “affinity diagram and image-schematic metaphors” (M = 3.45, SD = 1.21). However, the differences were not significant, $F(2,32) = 2.35$, $p = .11$, $\eta_p^2 = .13$. The ratings of the prototypes did not differ between conditions neither regarding the questions of how ”intuitive to use” (“affinity diagram only”: M = 5.2, SD = 1.5; “affinity diagram and image-schematic metaphors”: M = 5.5, SD = 1.0; “image-schematic metaphors only”: M = 5.4, SD = 1.2) or ”age-inclusive” (“affinity diagram only”: M = 4.7, SD = 1.7; “affinity diagram and image-schematic metaphors”: M = 5.4, SD = 1.1; “image-schematic metaphors only”: M = 5.0, SD = 1.4) the prototype would be later on.

Qualitative feedback on the method of image-schematic metaphors. Participants of the two conditions including image-schematic metaphors were asked to make clear perceived advantages and disadvantages of image-schematic metaphors in the design process. Feedback on the method of image-schematic metaphors could not be collected in the condition “affinity diagram only”. In general, three groups of feedback were collected: positive feedback, negative feedback and concrete recommendations on improvements of the methodology.

In the category of *positive feedback*, both conditions “affinity diagram & image-schematic metaphors” and “image-schematic metaphors only” received similar comments, which were grouped into three clusters. First, both methods were described as providing a basic structure and focus while leaving enough space for ideas and diversion (e.g., the image-schema CONTAINER). Second, participants perceived both conditions as stimulating creativity by providing a basis for brainstorming design ideas and fostering interaction and discussion between team members. Third, both conditions were described as easy to understand and apply.

In the category of *negative feedback*, differences between the conditions “affinity diagram & image-schematic metaphors” and “image-schematic metaphors only” were found, resulting in four clusters. First, some image-schematic metaphors in both conditions were seen as too abstract, inhibiting the understanding of each metaphor (e.g. TRANSACTION IS COMPULSION). One participant also criticised the English form of image schemas. Second, participants criticised the substantial amount of useless ideas that image-schematic metaphors stimulated (e.g. different instantiations for the image schema CONTAINER like drawers or boxes that were discarded finally). Third, only in the condition of “affinity diagram & image-schematic metaphors”, participants reported that the large list of image-schematic metaphors had limited their creativity and design process unnecessarily. Fourth, only in the condition of “image-schematic metaphors only”, minor concerns about the methodology were stated. For example, two participants out of twelve (from different design teams) described the provided time for the Wall Walk as too short. One participant argued that this methodology might be suitable for a team size of two, but would become difficult for scaling team size.

The category of *recommendations for improvement* consisted of three clusters. First, the presentation of image-schematic metaphors was requested to be more concrete in both conditions and – if possible – in the participants' native language German. Second, only in the condition "affinity diagram & image-schematic metaphors", a broader variety of image-schematic metaphors was favoured over the selection used in this study. Third, major or minor changes in the methodology arose for both conditions. For example, one participant stated that the affinity diagram and the image-schematic metaphors were not well enough connected and should be interwoven in the design process and wall walk more deeply. In the condition "affinity diagram & image-schematic metaphors", the small team size of only two participants was seen as positive and participants recommended refraining from using the method with larger teams. Also, participants suggested removing obviously useless or redundant ideas already in the earliest possible stage of the design process.

6.2.4 Discussion

Study 3 systematically investigated the method of image-schematic metaphors from the interaction designers' perspective. Based on the literature (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013), the ideation phase of Contextual Design – the Wall Walk – was enhanced with image-schematic metaphors. Participants completed a Wall Walk in three conditions: "affinity diagram only", "affinity diagram and image-schematic metaphors" and "image-schematic metaphors only". It was expected that the method of image-schematic metaphors would be as applicable as a standard method like the affinity diagram (*H1*) on different subjective characteristics of the design process like comprehensiveness, experience of flow, teamwork, emotions and mental workload (*H2*). Also, it was expected that participants would perceive the ideation process with image-schematic metaphors as more creativity stimulating than with the affinity diagram, generate more design ideas during the Wall Walk with image-schematic metaphors (*H3*) and rate their created prototypes as more innovative in conditions with image-schematic metaphors (*H4*).

Both conditions with image-schematic metaphors were rated as applicable as the standard design method (affinity diagram), supporting *H1*. Qualitative feedback also revealed that participants perceived the method of image-schematic metaphors as easy to understand and interesting in its applications. However, the introduction of image-schematic metaphors in the design process was not always beneficial. The combination of "affinity diagram & image-schematic metaphors" was descriptively rated lowest on almost all scales. The condition "affinity diagram & image-schematic metaphors" was descriptively perceived as providing worse teamwork, less structure and less inspiration for creativity. However, no significant difference was found in the quantitative measurement. Even though the cognitive load (SMEQ) showed a significant increase over the four phases of the experiment, the design condition did not influence the cognitive load. No differences were found in the measures of the experience of flow, emotions (PANAS), and perceived workload (NASA-TLX), which partially supports *H2*.

Participants generated significantly more design ideas with the image-schematic metaphors Wall Walk ("image-schematic metaphors only") than in the conditions "affinity diagram only" and "affinity diagram & image-schematic metaphors", which supports *H3*. However, participants did not perceive their prototypes as more innovative in conditions with image-schematic metaphors. Therefore, *H4* must be discarded. Especially the last finding is not in line with predictions from the literature. However, designing with image-schematic metaphors should increase the perceived innovation of the final users. In chapter 7, the paper prototypes of study 3 were developed to digital and interactive prototypes. Interaction designers do not rate their prototypes as more innovative when designing with image-schematic metaphors. But at this point, the evaluation of the final prototypes of study 3 is still due and will be the focus of study 6.

The qualitative feedback on the design process provided useful insights for improvements for the method of image-schematic metaphors. On the positive side, image-schematic metaphors are in fact as an applicable method to stimulate creativity. Participants also stated that most image-schematic metaphors are easy to understand and lead to novel ideas. The high applicability replicate previous findings (Hurtienne, Klöckner et al., 2015; Löffler et al., 2014) with larger sample size. Still, participants see the set of sometimes too abstract metaphors as a severe constraint and an essential limitation of the method. Biskjaer et al. (2010) also stated that metaphorical design tends to focus on only a few sources of inspiration which can constrain the design process concerning achieved stimulation for creativity. Also, many useless design ideas hamper the efficiency of the method of image-schematic metaphors and could be removed already in early phases of the design process. Löffler, Hess, Maier et al. (2013) also reported this overhead of pointless design ideas that are inspired by image-schematic metaphors. Finally, the collected qualitative feedback and given recommendations provide a reasonable basis for further developing and improving the method of image-schematic metaphors.

Some limitations must be considered when interpreting the results of this study. First, the study followed a mixed methods approach that combined qualitative as well as quantitative data (Trotter, 2012). The quantitative differences found in this study between conditions were small and much more extensive sample sizes would be required for statistical significance of the differences (Faul, Erdfelder, Lang & Buchner, 2007). However, the sample size was still large enough to provide some confidence in the descriptively found differences between conditions. Additionally, the mixture of quantitative as well as qualitative data allowed for verifying that participants did not differ systematically between conditions regarding possibly relevant characteristics such as prior experience or a priori knowledge about online-banking.

Second, the design process was challenging to standardise, which might have an impact on the internal validity of the study. Even though the study included only selected phases of the design process and controlled the procedure and setting over conditions as much as possible and feasible in this context, the high variance in the quantitative data shows that the procedure of the study was not comparable to a wholly controlled experimental study. Given the focus of the study and the research question, a complete standardisation was not feasible in this context without putting external validity in danger. However, since participants were not professional interaction designers but students in HCI-related degree courses, also the external validity of the study must be considered. Even though only this population of participants allowed for larger sample size, further studies should also address the question, how professional interaction designers perceive the method of image-schematic metaphors similar to smaller investigations such as (Hurtienne, 2011; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013; Löffler, Hess, Maier et al., 2013).

To conclude, study 3 showed that the method of image-schematic metaphors is rated as applicable as a standard method like an affinity diagram. This replicates the finding of Hurtienne, Klöckner et al. (2015) who underlined the straightforward application of this method. Method for interaction design that can stimulate many design ideas. However, integrating it into existing design methods like an affinity diagram can lead to adverse effects like poor teamwork. Reasons for this might be that the size of an affinity diagram might have drawn too much attention of the participants or that the two methods trigger different design strategies (e.g. diversion vs conversion Biskjaer et al., 2010). Thus, the recommendation of using image-schematic metaphors as a method is to use it as a separate phase in the design process (for example, by conducting a separate Wall Walk for brainstorming design ideas only with image-schematic metaphors instead of combining both methods).

6.3 Study 4: Image-Schematic Metaphors for Professional Interaction Designers

The standardised setting of study 3 allowed for a – not large enough for statistic significance – but still more extensive than the usual sample size. However, participants were restricted to students with only little professional experience. Also, the design process was not comparable to working circumstances in a design agency. The underlying question of study 4 is therefore on how the method of image-schematic metaphors blends into a design agency with professional interaction designers. Despite previous efforts to get insights about best practices for introducing image-schematic metaphors into professional design processes (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013), knowledge about the perceived potential for interaction design under realistic circumstances is limited. Study 4 contributes to the literature by providing insights from blending the method of image-schematic metaphors into the project-work of an established interaction design agency.

An experimental setup is difficult to achieve in a realistic context, and empirical studies on the design process are not comparable to psychological research (Zimmerman et al., 2007). However, qualitative feedback from professional interaction designers is needed to validate the insights obtained in study 3 with students. Especially the focus on professional interaction designers is a novel contribution to the literature. Despite the explorative approach of study 4, the expectations are similar to those in study 3: interaction designers should describe the method of image-schematic metaphors as stimulating creativity while at the same time providing structure to the design process.

6.3.1 Method

The study was conducted in cooperation with an agency with approximately 35 to 40 employees specialised in interaction design, user experience design, and app development. The project context aimed at the development of a solution that brings together different streaming services such as Netflix, Amazon Prime, and online media libraries of public channels on a smartphone. The method of image-schematic metaphors enhanced two phases of the project: a) user research for elicitation of image-schematic metaphors and b) interaction design for applying image-schematic metaphors in the ideation and prototyping phase.

User Research Seven young users of streaming services participated in contextual interviews. Only five of them gave their permission to audio-taping and transcribing the interviews. The utterances of these five interviews served as the basis for the extraction of image-schematic metaphors. The mean age of participants was 22.2 years (SD = 4.3), and two participants were female. The project scope itself did not allow the inclusion of older adults during user research.

During the contextual interviews, participants completed a variety of tasks with current streaming services (e.g., find information of a specific series in Netflix and save it to watch it later on). In line with the methodology suggested by Löffler, Hess, Hurtienne et al. (2013), participants were asked to think aloud to collect enough user utterances for language analysis. Two extractors (a student with extraction experience from a previous project and the author of this work) together analysed the transcribed interviews for image schemas and image-schematic metaphors. Importantly, extractors ignored image-schematic metaphors that were not relevant at this stage of the project (as a lesson from study 2). The final list contained 23 different image-schematic metaphors that stemmed from 82 user utterances. The extractors reduced this list to eight image-schematic metaphors on account of a) the number of instantiations in the user utterances and b) the relevance for the first design sprint (addressing the functionality of the first design sprint). Two image-schematic metaphors remained on the list despite their low frequency because they addressed core elements of the user interface. Table 6.5 summarises the final set of image-schematic metaphors.

Table 6.5: Image-schematic metaphors extracted for study 4.

Metaphor	Example for user utterance	Number of instantiations
VIDEO SELECTION PROCESS IS PATH	"It starts with movies I can watch, *then* my list,. . . "	16
CATEGORY IS CONTAINER	"they are grouped *in* categories"	7
ADDING TO WATCHLIST IS IN	"I push that *into* my watchlist"	5
IMPORTANCE IS BIG	"my wishlist has a *big priority* for me"	3
ADDITIONAL INFORMATION IS SUPERIMPOSITION	"On *mouseover* I get the information"	2
FRIEND FEED IS MATCHING	"friends recommend and that *matches* my favorites"	2
ADDING TO WATCHLIST IS COMPULSION	"I *push* that into my watchlist"	1
PAYMENT REQUEST IS BLOCKAGE[1]	"I can *only* watch this *when* I pay"	1

Workshops - Sample Two one-day workshops were conducted in the design agency, one with and one without the image-schematic metaphors. Per workshop, two professional interaction designers attended (all male, different interaction designers per workshop). Additionally, a senior interaction designer (not familiar with image-schematic metaphors) was ready for consultation throughout the design process but did only provide high-level feedback on original concepts. The experimenter moderated the method and schedule but did not intervene in the design process per se (except the introduction of the method of image-schematic metaphors). All four participants of the workshops rated their expertise in interaction design in general and UI/UX-design specifically on a 7-point Likert scale. Based on their ratings, the two design teams were assembled in a balanced way regarding expertise (see table 6.6). Participants did not receive a gratification since the workshops fell in their working time that their company paid.

Table 6.6: Description of the four participants in study 4. UI: User Interface. UX: User Experience. All participants usually worked together in larger project teams.

	Non-IS-M Workshop		IS-M Workshop	
	P1	**P2**	**P3**	**P4**
Age (all male)	22	28	17	20
Working years as interaction designer	2.0	3.5	1.5	3.0
Experience – interaction design (Novice: 1 to Expert: 7)	3	7	3	5
Experience – UI/UX-design (Novice: 1 to Expert: 7)	6	7	5	7

Workshops - Procedure Both workshops were conducted on one of two successive days in the same meeting room of the design agency. Timeslots of 6 hours were scheduled. The two design teams worked independently and did not communicate about the project. In both conditions, participants were instructed to follow their usual design procedure and work similarly to the way they would usually do in a real project. The workshops were part of the real project and results, and design concepts of both conditions were fed into the real project later on. Thus, participants were aware that their time in the workshops was valuable for the project itself.

In both conditions, participants gave informed consent at the beginning of the work-

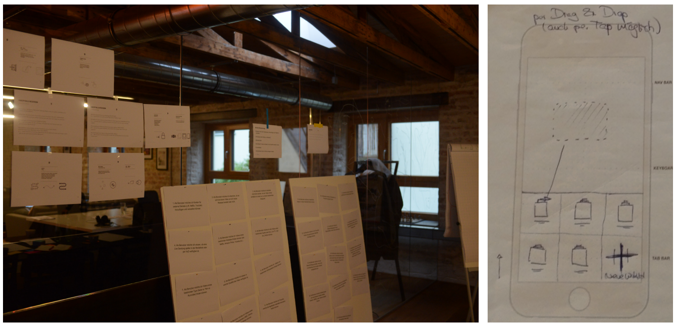


Figure 6.2: Setting of study 4 and exemplary concept of one screen of the paper prototypes.

shop. Based on the contextual interviews, 32 user stories were provided during the design process to ensure comparable functionality for the concepts in both workshops. Each user story consisted of a short description of a typical problem (e.g. "As a user, I want to find videos of a specific source like cinema, Live TV etc.") that was comparable to higher level notes of an affinity diagram used in Contextual Design (Holtzblatt & Beyer, 2017). The user stories were printed on cardboard and pinned to a wall. Also, participants had enough space and material for traditional paper prototyping and sketching (Holtzblatt & Beyer, 2017). The two design teams received identical information. The experimenter did not interact with the participants except for the introduction and overall organisation of the workshop. The design process of the two workshops deviated slightly and will be described separately in the following.

The team without image-schematic metaphors (the "Non-IS-M Team") started their – self-chosen – design process with a short analysis of similar applications (e.g. Netflix, Spotify) before examining the results of the user research. Afterwards, the team tried to translate abstract solutions from the user stories into concrete features for the following concept. For example, the user story "I want to see directly whether a video is a new release or not" was translated into a feature for the home screen. Translating all user stories to a screen or feature took approximately two hours. Afterwards, they defined main screens with their features (e.g. home screen, profile) and collected design ideas for each feature. Based on this developed information architecture, the design team discussed the result and started focusing on visual characteristics of single design elements.

The procedure of the team with image-schematic metaphors (the "IS-M Team") started with an introduction to the method of image-schematic metaphors. This introduction included the theoretical foundations and the overall process, examples for image-schematic metaphors as well as their application in user interfaces, and the outline for the workshop. Image-schematic metaphors from table 6.5 and one exemplary user utterance were also printed on cardboard and pinned to the wall. The focus was on providing examples of image-schematic metaphors that would later also be relevant in the design phase. The explanation was based on materials provided by Löffler, Hess, Hurtienne et al. (2013). Participants were not forced to use all image-schematic metaphors in the final concept but should keep them in mind nevertheless. In total, the introduction took 40 minutes. As soon as participants had no further questions, they moved into the conception phase. Even though the "IS-M Team" began by briefly analysing the image-schematic metaphors, the structure of the design process was comparable to the "Non-IS-M Team".

The user stories were used as the foundation for the information architecture and the basic structure of the application. Concrete design ideas were explored, and the team discussed the overall structure of the application and concrete user interface elements. Finally, the team also addressed essential visual characteristics of the prototype.

After the design phase, each participant was separately interviewed approximately 10 minutes about the experience during the design process and the used methods. During the interview, the focus was on the effectivity and efficiency during the design process. To better compare qualitative results, participants rated on Likert-scales the quality of the teamwork and the difficulty of ideation during the design process.

6.3.2 Results

Due to the small sample size, the focus of the study was on qualitative feedback during the design process. However, the collected quantitative data are at least descriptively in line with the expectations (see table 6.7). Participants in the "Non-IS-M Team" perceived generating ideas as more difficult compared to the "IS-M Team". Also, one participant in the "Non-IS-M Team" rated the teamwork as inefficient. The rating of the support by the method of image-schematic metaphors was in the scale mean.

The qualitative interviews after the design session (focusing on the "IS-M Team") and the created concepts of the two design teams provide more insights into the perception of the method by professional interaction designers and the applicability of the method.

Table 6.7: Quantitative ratings in study 4.

	Non-IS-M Team		IS-M Team	
	P1	**P2**	**P3**	**P4**
Difficulty of generating design ideas (Easy: 1 to Difficult: 10)	6	6	3	3
Efficiency of teamwork (Not Efficient: 1 to Efficient: 10)	9	3	8	9
Support by the method of IS-M (No Support: 1 to Very Supporting: 10)	/	/	6	5

Compatibility with the company's standard design process The company usually adopts a standard design process, which includes user research as well as interaction design. Interestingly, the method of image-schematic metaphors was similar to particular characteristics of the company's methodology. Specifically, the company uses an approach that builds on so-called user experience identities (e.g., SAFETY, EASE) the final user interface is supposed to convey (van de Sand, 2017).

These identities are an essential part of the design process, and designers use them both as a source of inspiration during the interaction design as well as a basis for final discussions of the resulting interface. Interestingly, the approach of identities can be regarded as having a similar function like image-schematic metaphors but on a different level. While identities play an essential role for high-level design decisions such as style guides (e.g., colour schemas, rounded vs sharp corners), image-schematic metaphors can be seen as working on single interaction-elements or user interactions (e.g., drag-and-drop, layout). The approach of identities argues – also similarly to image-schemas – that people learn signs for specific identities during the first seven years of childhood, e.g., round shapes are associated with OPENNESS and HUMAN (van de Sand, 2017). In sum, the method of user experience identities can be regarded as harmonious with the method of image-schematic metaphors.

Qualitative feedback during the design process Even though the team did not work over a more extended phase exclusively with the image-schematic metaphors, they directly affected the design process in all phases. For example, image-schematic metaphors were incorporated into the communication between designers ("this is a container", "we still need the blockage here") and considered carefully during the conception of the layout of the single screens. However, when an image-schematic metaphor was not seen as valuable or interfered with the chosen concept, it was directly discarded. Both participants in the "IS-M Team" stated that the method of image-schematic metaphors should not replace other methods in the design process and that – as in the current study – integrating them into the procedure by choice was the best approach. Notably, the company's method of user experience identities should be considered when adding new design methods like image-schematic metaphors.

Participants saw image-schematic metaphors as a useful basis for ideation, but they were sceptical about incorporating all of them in one user interface. They underlined the value of deliberately designing against single image-schematic metaphors as a source of inspiration. One important topic was the compatibility of image-schematic metaphors with existing guidelines and design patterns, especially in the context of mobile applications. Here, designers often have to apply a variety of recommendations and principles. Participants were concerned about following only image-schematic metaphors instead of, for example, principles of consistency.

Participants did not perceive the method of image-schematic metaphors as providing additional structure to the design process compared to the standard process including user experience identities. However, one participant admitted that the communication had changed during the concept phase compared to the usual procedure in similar projects. In the beginning, the two team members did not talk about concrete design solutions but used the abstract terminology of the image-schematic metaphors to refer to abstract interaction elements. For example, one participant described the abstractness of the image-schematic metaphors as positive because they forced him to go back one step in the design phase and not stick too early to concrete design solutions. The other participant found them too abstract and would prefer more concrete examples in the form of "compulsion is drag-and-drop". Especially for dynamic content such as gestures and animations, a high potential was seen in the method of image-schematic metaphors that could support the design process on interaction elements that are otherwise difficult to communicate in a transparent manner.

In sum, the method of image-schematic metaphors did not completely change the company's standard methodology and design processes. Designers still followed their routine similarly than in the baseline, but one participant remarked that design decisions were made less spontaneously than usual. The majority of design decisions was in accordance with relevant image-schematic metaphors. Also, both participants saw the quotes of participants for each image-schematic metaphor as very useful to make them less abstract.

Image-schematic metaphors in the final concepts One team worked without and the other with image-schematic metaphors, which allowed a comparison between their two resulting concepts.[2] Even though the "Non-IS-M Team" did not receive the explicit image-schematic metaphors, their concept was not free of image-schematic metaphors. In sum, the "IS-M Team incorporated seven out of eight provided image-schematic metaphors. At the same time, the "Non-IS-M Team" also incorporated four out of eight image-schematic metaphors into their user interface. For example, IMPORTANCE IS BIG was instantiated in the form of a (compared to the other interaction elements) enlarged "Highlight-Button" for important new videos.

In the"IS-M Team", participants easily translated the image-schematic metaphors with a CONTAINER into concrete design ideas. Here, the instantiation in the user interface was mainly a frame around different categories with a semantically relevant in- and outside. Furthermore, the COMPULSION image schema was also easy to apply in the form of a drag-and-drop gesture.

The set of image-schematic metaphors provided designers with specific prioritisation for user stories and interface elements. For example, the image-schematic metaphor CATEGORY IS CONTAINER made visible the importance of different categories on a top-level. Also, ADDING TO WATCHLIST IS IN gave the impulse for placing a permanently visible button for the watchlist on most screens instead of hiding this feature in a sub-menu as happened in the "Non-IS-M Team".

Lastly, the collection of image-schematic metaphors provided a form of a checklist for the created concept, and the design team went through all of them while checking if the interface was in line with them. The image-schematic metaphor IMPORTANCE IS BIG was not integrated into the concept, which was recognised in the final discussion. Still, integrating it would have required significant changes in the concept and was therefore discarded.

[2]Due to nondisclosure, only essential and methodology-relevant characteristics of the final designs are summarised here.

Table 6.8: Image-schematic metaphors and their integration into the concept phase. In the "IS-M Team", the metaphors FRIEND FEED IS MATCHING *and* PAYMENT REQUEST IS BLOCKAGE *were initially discussed, but incorporated very late in the design process.*

Metaphor	Non-IS-M Team	IS-M Team
VIDEO SELECTION PROCESS IS PATH	✓	✓
CATEGORY IS CONTAINER	χ	✓
ADDING TO WATCHLIST IS IN	χ	✓
IMPORTANCE IS BIG	✓	χ
ADDITIONAL INFORMATION IS SUPERIMPOSITION	✓	✓
FRIEND FEED IS MATCHING	χ	✓
ADDING TO WATCHLIST IS COMPULSION	χ	✓
PAYMENT REQUEST IS BLOCKAGE	✓	✓

6.3.3 Discussion

In this study, the method of image-schematic metaphors was integrated into an existing workflow of an interaction design agency. The aim was to collect insights on a) obstacles and facilitators for the introduction of the method and b) the effectiveness of image-schematic metaphors in the design process. In two interaction design workshops, professional designers created a concept for a real project either without or with image-schematic metaphors that had been extracted from user research in the same project. Due to the expensive and therefore small sample, the focus was on replicating and validating in-the-wild the qualitative results obtained in study 3. Specifically, it was of interest whether the method was applicable and creativity-stimulating for the interaction designer and whether interaction designers would incorporate instances of image-schematic metaphors into the user interface. One critical difference of this study to previous work (Löffler, Hess, Maier et al., 2013) is the focus on professional interaction designers as a vital and highly relevant target group for proliferating the method of image-schematic metaphors into praxis.

A first analysis of the standard design process of the company (user experience identities van de Sand, 2017) regarding its compatibility with the method of image-schematic metaphors showed that both methods share some basic foundations (e.g., grounding the interface on fundamental principles and experiences learned in the childhood). Since both approaches focus on different levels of interaction design (user experience identities: high-level; image-schematic metaphors: low-level), they could be complementary in the interaction design process.

In line with study 3, interaction designers perceived the method of image-schematic metaphors as comprehensive and easy to apply, which also supports other previous findings (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013). The six hours of the workshop were sufficient to convey the underlying concepts of the methodology and theory as well as doing the first iteration of interaction design. However, participants descriptively described the method as neither constraining but also not as very supporting in the design process. This is in also contrast a) the findings of study 3 and b) to the observation that participants, in fact, adopted the terminology of image-schematic metaphors during the design process. Also, participants of the "IS-M Team" stated that image-schematic metaphors helped in communicating on a conceptual level instead of directly on the detailed user interface level.

One explanation for the decent appraisal of the method of image-schematic metaphors could be that professional interaction designers followed mostly their standard design process that was also observed in the "Non-IS-M Team" and the introduction of image-schematic metaphors did not lead to major changes. Image-schematic metaphors were not forced into the design process, which could have affected the success of the method in the company's standard procedure neither positively nor negatively. Still, the "IS-M Team" emphasised that the little intrusive addition of image-schematic metaphors would be the right approach. Also, pressure on interaction designers to use image-schematic metaphors should not be increased in the future.

Both quantitative, as well as qualitative data, indicated that participants perceived image-schematic metaphors as stimulating creativity. The ratings for the difficulty of finding new ideas during the design process were descriptively lower for the "IS-M Team" compared to the "Non-IS-M Team". Also, participants used image-schematic metaphors on several occasions to return from a dead end in the design process and sought in them a source of inspiration. During the interviews after the design workshops, it became clear that interaction designers saw the abstractness of the image-schematic metaphors as both an advantage and a disadvantage: image-schematic metaphors provide rough guidance without constraining their creativity but are also not concrete enough to have a particular impact on design decisions compared to design guidelines.

The analysis of the concepts that resulted from the design workshops revealed that the "IS-M Team" did not incorporate all image-schematic metaphors in the user interface and left out one of them entirely. Two image-schematic metaphors were added very late in the design process. In sum, only five out of eight image-schematic metaphors were used initially as a basis for the concept. One explanation for this could be that interaction designers regarded the single image-schematic metaphors as differently comprehensible. For example, the image-schematic METAPHOR WATCHLIST IS CONTAINER was more natural to translate into a specific feature in the user interface than FRIEND FEED IS MATCHING. Differences in comprehensibility between single image schemas have been reported (Hurtienne, 2011) and the CONTAINER image schemas seems to be one of the most reasonable ones also during the extraction process from natural language (Winkler et al., 2016). Also, the image schema COMPULSION was directly associated with drag-and-drop in the graphical user interface, raising the question if the terminology of image schemas must be kept in real projects or if it would benefit from re-formulation for better comprehensibility. Similar to the confusion during the extraction process from natural language, the translation of image schemas to concrete design decisions must be considered when practising this method. Many examples of image schemas, image-schematic metaphors, and their application in HCI have been reported – for example in the ISCAT-database (Hurtienne, 2017a). However, if this information is not accessible and practical enough for application in real projects this information is in danger of facing the same problems as interface guidelines, which exist but are not frequently used (Dickinson et al., 2007).

On the other hand, the "Non-IS-M Team" integrated four out of eight metaphors without instruction. This is an interesting finding since user interfaces will mostly incorporate image schemas and image-schematic metaphors even when the method of image-schematic metaphors is not explicitly known to the interaction designer. However, a list of image-schematic metaphors is needed to guide interaction designers to incorporate at least most of the "right" ones.

Due to several limitations in the context of a real company, study 4 is a form of case-study. First, the cost-intensive participants did not allow a more extensive study with more interaction designers. Thus, the small sample size and only one design process with image-schematic metaphors render the few quantitative results as difficult to interpret. The qualitative feedback must not be generalised without keeping in mind these limitations, even though qualitative insights from two complete one-day workshops were achieved. However, the focus of this study was apparently on qualitative feedback and observations. Also, the quantitative measures are difficult to interpret given the small sample size.

Second, even though two design teams were available for the study, it was decided to let only one of them work with image-schematic metaphors. On the one hand, this eliminated the possibility of comparing two design process with image-schematic metaphors. On the other hand, only this approach allowed for comparing the design process enhanced with image-schematic metaphors to the standard process with user experience identities. In fact, the "IS-M Team" did not deviate much from their usual procedure. This is an insight that would not have been possible without the baseline "Non-IS-M Team".

Taken together, study 4 shows that image-schematic metaphors can blend into the processes of a professional design procedure. Even though this study did not focus on older adults and age-inclusiveness, these insights are valuable and a pre-requisite for progressing the method of image-schematic metaphors for professional interaction design. Importantly, the method of image-schematic metaphors is readily applicable for interaction designers.

6.4 Summary

Previous work already showed positive feedback from project teams on the method of image-schematic metaphors (Hurtienne, 2011; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013). However, research gaps were identified in the literature. Importantly, some studies did not include baseline design processes (e.g. Hurtienne, Klöckner et al., 2015), rendering the interpretation of the provided feedback and ratings difficult.

This chapter described two studies that focused on the designer's perspective. Compared to a baseline design method (study 3: affinity diagram; study 4: the company's standard design procedure), it was expected that participants would perceive the method of image-schematic metaphors as equally applicable. Also, it was expected that participants would perceive the method of image-schematic metaphors as creativity stimulating and leading to innovative design solutions.

Study 3 applied the method of image-schematic metaphors using three conditions in an experimental setting. In study 4, one team of professional interaction designers created a concept for a real project following a standard design process (User Experience Identities, van de Sand, 2017), while another team worked with the standard design process which was enhanced with image-schematic metaphors. Both studies conclude that the underlying concept of image-schematic metaphors is easily applicable for novices and professionals. While novices in study 3 described the method of image-schematic metaphors as creativity stimulating, professionals in study 4 were more reluctant in seeing image-schematic metaphors as a tool for creativity and innovation. Still, novices and professional designers underlined the fact that image-schematic metaphors provide general guidance for interaction design that at the same time ensures designing in line with the mental model of potential users as well as allowing for spontaneous ideas.

In studies 3 and 4, image-schematic metaphors were integrated into two different standard design processes. Given the results of both studies, image-schematic metaphors were difficult to integrate into the procedure of affinity diagram but easy to integrate into the process of User Experience Identities and User Stories. One explanation for this could be that professionals in study 4 were more experienced than the novices from study 3. It might also be that an affinity diagram was less compatible with image-schematic metaphors, even though (Hurtienne, Klöckner et al., 2015) report good compatibility between both methods. However, image-schematic metaphors, used as a supplementary method supporting a primary design method, should be considered carefully before integrating them into the other design procedure. In sum, the two studies also revealed several insights that can help to improve the methodology for applying image-schematic metaphors in practice:

Separation of standard procedure and image-schematic metaphors Using image-schematic metaphors as a supplementary method supporting a standard design method should be considered. The design team should perform a separate Wall Walk with image-schematic metaphors alone instead of integrating image-schematic metaphors into an affinity diagram.

Applicable method for interaction designers Participants in both studies comprehended the general concept of designing image-schematic metaphors after an introduction of 30 to 45 minutes. Professional interaction designers were able to integrate image-schematic metaphors into specific phases of their standard design process (especially at the beginning and the end of the workshop).

Examples and Quotes Participants perceived generating concrete instantiations of image-schematic metaphors in the user interface as challenging. Thus, designers should be provided with examples of user utterances that led to the extraction of single metaphors and examples for possible instantiations of single metaphors in user interfaces. However, difficulties can also represent a source of inspiration and stimulate out-of-the-box thinking (as stated in study 3). Minor challenges when thinking about proper instantiations of a specific image-schematic metaphor might therefore even be beneficial for the design process.

Until now, the theoretical foundation for innovative and age-inclusive interfaces has been laid, and challenges for incorporating image-schematic metaphors in the design process have been investigated. Still, the final question for this thesis remains: can image-schematic metaphors lead to innovative and age-inclusive interfaces? The next chapter addresses this question describing two more user studies that evaluated interactive prototypes regarding their perceived innovation and age-inclusiveness.

Chapter 7

Innovation and Age-Inclusiveness of IS-M Prototypes

Chapter 5 showed that younger and older adults share a substantial part of their mental models of abstract concepts, operationalised via image-schematic metaphors. Chapter 6 provided recommendations on how to integrate image-schematic metaphors into the interaction design process as a source of inspiration and stimulation for innovative solutions. Both chapters taken together lay the foundation for testing for age-inclusiveness and innovation simultaneously. While previous work supports the feasibility of designing for innovation and age-inclusiveness at the same time (see chapter 4), evaluations of interactive prototypes and empirical data stemming from user tests are still rare. Moreover, comparing prototypes complying with image-schematic metaphors to baselines is often missing (Hurtienne, Klöckner et al., 2015).

The presumed advantage of image-schematic metaphors is that they can directly address the tension between tradition vs. transcendence (building for the current status-quo vs. creating a novel solution). These two goals are often seen as being diametrical to each other (Biskjaer et al., 2010; Ehn, 1988). Hurtienne, Klöckner et al. (2015) already provide first empirical data on resolving this tension. However, due to a missing baseline and only comparing younger and older adults by age and not by cognitive abilities and prior technological knowledge, the positive and promising results need to be interpreted with caution. The work reported by Winkler et al. (2016) compared image-schematic metaphors-prototypes to baselines, but due to a small sample size, the generalisability of the findings is limited.

This chapter describes two studies that compared prototypes explicitly designed to comply with relevant image-schematic metaphors (IS-M prototypes) with a baseline (industry standard or baseline design process, Non-IS-M prototypes). These user studies with younger and older adults focused, on the one hand, on the users' perceived innovation of the prototypes. On the other hand, the age-inclusiveness was evaluated by comparing differences (performance and subjective measures while solving use cases) between both age groups for IS-M prototypes and Non-IS-M prototypes. In both studies, participants completed tests on their cognitive abilities and prior technological knowledge. These measures allowed for directly detecting two age-related changes that are relevant for HCI (instead of investigating extreme groups under the assumption that they would differ in cognitive abilities and prior technological knowledge). The here reported work allows – the first time – to compare a multitude of image-schematic metaphor prototypes to baselines and with larger participant samples.

Contrary to Hurtienne, Klöckner et al. (2015), this work focusses on innovation and age-inclusiveness and less on the promise for intuitive use. Although studies 5 and 6 slightly differ in their methodology, the same three hypotheses apply, which are outlined in the following. First, image-schematic metaphors are a technique for stimulating creativity and leading to innovative user interfaces:

H1 Prototypes that were designed to explicitly based on image-schematic metaphors are perceived as more innovative by participants than prototypes that were designed without image-schematic metaphors.

Second, instead of measuring only chronological age as a proxy for age-related changes, the more direct measures of cognitive abilities and prior technological knowledge (Czaja et al., 2006; Kuerbis et al., 2017; Mitzner et al., 2010) are used to estimate the impact of age-related differences on the interaction with IS-M prototypes vs. Non-IS-M prototypes:

H2 Younger and older adults differ in their cognitive resources as well as in their prior technological knowledge.

Third, these age-related differences between participants should have less impact on performance measures (Hurtienne, Klöckner et al., 2015), as grounding HCI on image-schematic metaphors exploits fundamental sensorimotoric knowledge (that is independent of experience with technology). This knowledge can, in turn, be activated subconsciously requiring a minimum of cognitive resources:

H3 User performance and the subjective workload is less affected by age-dependent differences (cognitive resources, prior technological knowledge) for prototypes that were designed using image-schematic metaphors compared to prototypes that were designed without image-schematic metaphors.

7.1 Study 5: Evaluation of Prototypes from Study 3 (Domain Online Banking)

In study 3, participants had created paper prototypes in one of three conditions. They were either supported by a) a classical Affinity Diagram, b) an Affinity Diagram and image-schematic metaphors or c) only image-schematic metaphors. For study 5, a visual designer (not the author of this work) evolved those paper prototypes into interactive digital prototypes. This was necessary to create a set of interactive prototypes that were designed explicitly with or without image-schematic metaphors resulting in twelve prototypes (four for each condition). The goal of the subsequent user test was to empirically compare the prototypes of different design teams in different design conditions in measures of innovation and age-inclusiveness.

7.1.1 Method

Sample In total, 47 participants were recruited for this study representing a bimodal age distribution. 24 of them were in the age group of younger adults (age: M = 22.8 years, SD = 2.7; 12 female) and 23 in the age group of older adults (age: M = 59.7 years, SD = 6.6; 10 female). These age-groups are similar to other studies related to age-differences (Sonderegger, Schmutz & Sauer, 2016). Importantly, the sample did not focus on chronological age itself as a proxy for age-differences but also directly measured differences in cognitive abilities and prior technological knowledge. Participants from the younger adults group were recruited via social media, personal networks and a local student panel of the university, while older adults were recruited via family members, senior residencies and private networks. Younger subjects received partial course credit for their participation and older adults participated voluntarily without compensation. Most younger adults had received Abitur ("Abitur": 23 participants, "University Degree": 1), while older adults differed more in their educational level ("Mittlere Reife": two participants, "Abitur": 10, "University Degree": 10, "PhD": 1).

Prototypes The prototypes in study 5 were taken from study 3 and covered two use cases in the domain of online banking. Because of the significant variance between the designed prototypes, not all of the original 18 paper prototypes were implemented into interactive prototypes. In a first walkthrough, six prototypes were excluded from further development based on two criteria: little compliance with basic usability principles (e.g. consistency, user control Nielsen, 1994) and missing functionality (e.g. missing login screen). A visual designer further developed the remaining 12 paper prototypes (four for each of the three design conditions of study 3) into interactive prototypes using Axure and Photoshop (see Appendix A. 4).

During this process, the layout and basic structure of the paper prototypes were strictly kept. A style guide ensured the standardisation of the visual appearance to a certain extent. The background colours were white (#FFFFFF) or a very dark blue (#003366). Text content was adopted mostly without changes. If modifications were made, they usually consisted of a change of the company's name to a standardised one ("Kontoverwalter"). The text font was set as "Arial", text size as 20 or larger, and text colour was either the white or dark blue.

Besides these colours for text, signal colours like red were used in some prototypes.

Most paper prototypes included some icons of varying quality. They were converted to vector graphics in Photoshop while keeping the underlying structure. To follow the colour scheme and to make interactions more visible, the colour of interactive elements such as buttons was consistently set to orange (#FF9900). To make interactive content more explicit in all prototypes, corners were rounded and shapes shaded (contrary to static content). To standardise the prototypes to a certain extent, necessary functionalities like a "back"-button" were added to each screen. Also, additional functions that were not part of the use case and increased the complexity of one prototype in comparison to the others were removed.

After the implementation, a short pluralistic walkthrough was conducted with two external testers familiar with usability engineering (blind to the conditions). This expert evaluation allowed revealing usability problems that were addressed before user testing. However, the underlying interaction design of the paper prototypes remained in its essence, but details that made interaction impossible or difficult were changed in all prototypes regardless of the condition (e.g. missing functionality to go one step back).

Experimental design Younger and older adults evaluated the interactive prototypes taken from the three design conditions of study 3, resulting in two factors for this study:

design condition (within factor, three levels) a) only Affinity Diagram, b) Affinity Diagram and image-schematic metaphors, and c) only image-schematic metaphors

age (between factor, two levels) a) younger adults and b) older adults

An incomplete mixed design was applied, and each participant interacted with six out of 12 possible prototypes. This approach was chosen because of three reasons. First, learning effects were expected to affect the results to a substantial extent for the number of twelve prototypes, and the duration of the experiment would have exceeded an acceptable study duration for older adults (Dickinson et al., 2007). Second, since the prototypes stemmed from three design conditions and each covered one of two use cases, each participant had to be confronted with at least six prototypes to allow a comparison over all conditions. Third, a complete between-subjects design with each participant completing only one or two prototypes would have increased the expected variance tremendously and lowered the data points for each condition significantly.

Participants tested the prototypes focusing on the following dependent variables: perceived innovation (AttrakDiff2-HQS), performance during the completion of the use cases (effectiveness: number of errors/deviation from the ideal path; efficiency: task time, mental effort) and the perceived intuitive use (QUESI). More details of the used instruments can be found in the following section.

Material Age was expected to be a proxy variable for differences in cognitive abilities and prior technological knowledge. Thus, participants in both age groups were tested regarding their cognitive abilities and prior technological knowledge. Cognitive abilities of the participants were measured using the following two instruments:

1. Trail Making Test (TMT) with part A and part B (see chapter 2). Participants have to connect either letters in their given order (TMT-A: *A-B-C*...) or alternatingly letters and digits to each other (TMT-B: *A-1-B-2-C-3*...). The time needed to complete one sheet full of letters or letters and digits is the primary score for this task. This test is a golden standard for measuring cognitive abilities, and especially the TMT-B has been linked closely to fluid intelligence and information processing speed (Rabin et al., 2005). By comparing the performance in TMT-A and TMT-B, further insights can be extracted from the task performance, because TMT-A can serve as a baseline for an easier task involving less cognitive resources but is similarly demanding visual search and psychomotor speed. The task performance has been shown to decline with age steadily, while error rates of the TMT-B are less sensitive to age-related differences in cognitive abilities, but can diagnose clinically relevant cognitive impairments (Ashendorf et al., 2008).

2. The Digit Letter Substitution Test. Participants are asked to translate a sheet full of letters with a given coding scheme to digits (e.g. “W” to “1”, “B” to “2”, “T” to “3”, see chapter 2). The number of substitutions in a given time-span serves as the main score for this tests. This task has been shown to validly measure information processing speed which declines with age (van der Elst et al., 2006).

Prior technological knowledge focusing on online banking was measured similarly to study 1 and was based on a questionnaire. However, due to time constraints of the experiment, the questionnaire was slightly reduced and did not include the Computer Literacy Scale (Sengpiel & Dittberner, 2008), because it was not sensitive to small differences in prior knowledge in study 1 and most participants had scored either very high or very low in study 1. Besides this alteration, all dimensions of Hurtienne et al. (2013) were part of the questionnaire (see chapter 5). As in study 1, a TA-EG questionnaire was applied, estimating the overall prior technological knowledge (Karrer et al., 2009). The measured facets of prior knowledge with technology were:

exposure - low specificity Participants reported how often they had used each of 15 different technologies in their life (e.g. a ticket vending machine, smartphone, personal computer; from ’never’ to ’regularly’; range of sum score: 15 to 75).

exposure - medium specificity Participants reported how often they had performed each of 20 different tasks with a screen-based technology (e.g. listening to music, social media, sending e-mails; ’never’ to ’daily’; range of sum score: 0 to 100).

exposure - high specificity Participants reported in form of a checklist which of 21 functions common in online banking they had ever used in their lives; range of sum score: 0 to 21).

competence - low specificity Participants completed a questionnaire on technology affinity (TA-EG; only the value of the subscale "competence" was used; range of average score: 1 to 5 (Karrer et al., 2009)).

competence - high specificity Participants subjectively rated their competence for online banking on a 7-Likert scale (range: -3 to +3 with a neutral "0").

From the collected data, two scores were computed for each participant: a) a cognitive abilities score and b) a prior technological knowledge score. The cognitive abilities score was the mean of the z-standardised values of the three measures (TMT-A, TMT-B, Digit Letter Substitution Test). The prior technological knowledge score was the mean of the z-standardised values of the five measures (three levels of technology exposure and two levels of technology competence).

To collect data on the perceived innovation of each prototype the subscale "hedonic quality – stimulation" of the AttrakDiff2 was applied (Hassenzahl, 2004). Finally, the QUESI served as an overall evaluation of the perceived intuitive use during the interaction with a prototype (Naumann & Hurtienne, 2010). It is a useful measurement of the subjective consequences of intuitive use and a standard tool in research focusing on intuitive use (Hurtienne, Klöckner et al., 2015).

Apparatus All user tests utilised a Microsoft Surface Pro 3 tablet with a diagonal of 12 inches (30.5cm) and a resolution of 2160 x 1440 pixel (pixel density 216ppi). The operating system was Windows 8.1 Professional. The luminance of the display (Clear Type-Display, Peak Brightness: 371cd/m2) was set to maximum, the tablet was laying on the table, and the stand angle was 30°.

Procedure Each session was conducted individually with one participant, and the location was mostly the university or, to make participation more convenient for older adults, their homes. After the introduction, an overview of the experiment and signing informed consent, each participant completed the questionnaire on demographic information and prior technological knowledge. Afterwards, the cognitive tests were applied (first: Trail Making Test; second: Digital Letter Substitution Test).

The central part of the experiment started with a short introduction to the tablet. Afterwards, participants interacted with six (out of the twelve possible) prototypes. Thus, each participant completed one prototype for each of the six possible combinations of design condition and use case. Prototypes were presented in pseudo-randomised order. After each prototype, participants completed the SMEQ-scale and the AttrakDiff2-HQS-scale. For the first encounter of each use case, participants additionally filled out a QUESI. The reason for administrating the QUESI only for two of the six prototypes were the strong time constraints of the experiment. At the end of the experiment, participants qualitatively described the prototypes they had evaluated during the experiment. The entire procedure took between 90 and 120 minutes.

7.1.2 Results

Data analysis was conducted using SPSS version 24, and the alpha-level was set to .05 if not stated otherwise. No participants were excluded, but single data were missing (e.g. due to technical problems or participants forgetting single items, 0.2% of all data points).

Perceived innovation Over all prototypes, both age groups differed in their perceived innovation (see figure 7.4). Results of an ANOVA with design condition and age group as independent variables and HQS as the dependent variable revealed also a significant influence of age-group on the perceived innovation with older adults rating prototypes as more innovative compared to younger adults, $F(1{,}274) = 16.71$, $p < .001$, $n_p^2 = .06$. Furthermore, a significant large effect was found for the design condition, $F(2{,}274) = 30.47$, $p < .001$, $n_p^2 = .18$, and prototypes that had been designed with image-schematic metaphors ("AD & IS-M" and "IS-M only") scored higher on perceived innovation than the baseline ("AD only"). The interaction between both factors was not significant, $F(2{,}274) = 2.39$, $p = .09$.

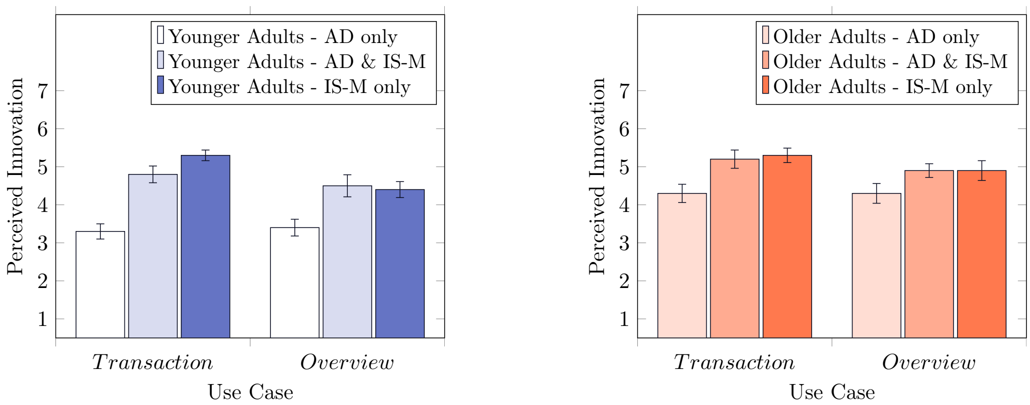


Figure 7.1: Mean and Standard Error of perceived innovation per age group, design condition and use case. AD: Affinity Diagram. IS-M: image-schematic metaphors. See also table 7.1.

Table 7.1: Perceived innovation per use case and design condition for both age groups. Mean & [SD]. AD: Affinity Diagram. IS-M: Image-Schematic Metaphors.

Use Case	Age Group	Younger Adults	Older Adults
Transaction	AD only	3.3 [1.0]	4.3 [1.1]
	AD & IS-M	4.3 [1.1]	5.2 [1.1]
	IS-M only	5.3 [0.7]	5.3 [0.9]
Overview	AD only	3.4 [1.1]	4.3 [1.3]
	AD & IS-M	4.5 [1.4]	4.9 [0.9]
	IS-M only	4.4 [1.0]	4.9 [1.2]

Cognitive abilities As expected, younger adults performed better in the cognitive tests compared to the age group of older adults. In the TMT-A, younger adults were significantly faster in completing the test (M = 23.5s, SD = 7.0s) compared to older adults (M = 36.4s, SD = 11.6s), $t(31.9) = 4.43$, $p < .001$, $d = 1.4$ (degrees of freedom adjusted due to variance inhomogeneity). Younger adults also completed the TMT-B significantly faster (M = 52.8s, SD = 23.1s) than older adults (M = 74.7s, SD = 35.8s), $t(37.3) = 2.49$, $p = .017$, $d = .73$ (degrees of freedom adjusted due to variance inhomogeneity). In the Digit Letter Substitution Test, younger adults substituted significantly more letters with digits ($M = 74.1$, $SD = 12.6$) compared to older adults ($M = 66.0$, $SD = 13.6$), $t(45) = 2.11$, $p = .040$, $d = .62$. Based on the three cognitive tests, an overall z-standardised cognitive score was calculated (younger adults: $M = .391$, $SD = .584$; older adults: $M = -.350$, $SD = .877$). For this sample, the score ranged from -2.21 to +1.58.

Prior technological knowledge Five dimensions of prior knowledge were measured. The sum score for technology exposition - low specificity (frequency of using each of 15 technology devices in general), was significantly higher for younger adults ($M = 45.4$, $SD = 5.5$) compared to older adults ($M = 37.9$, $SD = 8.7$), $t(45) = 3.54$, $p = .001$, $d = 1.03$. Younger adults reported to have more technology exposition - medium specificity (frequency of performing single 20 tasks with technological devices; $M = 58.7$, $SD = 10.7$) compared to older adults ($M = 40.6$, $SD = 15.8$), $t(38.2) = 4.60$, $p < .001$, $d = 1.3$. The age groups in this sample did not differ significantly in their technology exposition - high specificity (checklist of 21 functions already used in online banking; younger adults: $M = 5.92$, $SD = 4.56$; older adults: $M = 5.74$, $SD = 5.85$, $t(41.6) = .12$, $p = .91$).

Technology competence - low specificity (TA-EG questionnaire, subscale "competence") differed slightly between age groups: younger adults reported competence levels ($M = 3.41$, $SD = .54$) that differed marginally significantly from the group of older adults ($M = 3.15$, $SD = .43$), $t(45) = 1.79$, $p = .08$, $d = .53$. Technology competence - high specificity did not differ between age groups (self-reported competence in online-banking ranging from -3 to +3; younger adults: $M = 1.13$, $SD = 1.69$; older adults: $M = 1.13$, $SD = 1.96$, $t(44) = .00$, $p = 1$).

An overall prior knowledge-score was calculated which was z-standardised (younger adults: M = .258, SD = .559; older adults: M = -.267, SD = .731). For this sample, the score ranged from -1.80 to +1.38.

Completion time Over all prototypes, younger adults completed the use cases significantly faster than older adults (see table 7.2), $t(194.95) = 12.79$, $p < .001$, $d = 1.54$. Also, the completion times significantly differed between the design conditions, $U(2) = 15.51$, $p < .001$.[1] Bonferroni-adjusted pairwise comparisons showed significant differences between the conditions "AD only" and "AD & IS-M", $U(1) = 35.6$, $p = .008$, and "IS-M only" and "AD & IS-M", $U(1) = 3.7$, $p = .001$.

To test the hypothesis that cognitive abilities and prior knowledge should affect the completion time for the design condition "AD only" to a lesser degree than for the design conditions "AD & IS-M" and "IS-M only", three regressions were computed (one for each design condition). The explained variance was compared between the three models. The results of these regression analyses are summarised in table 7.3.

In sum, age as a proxy variable for several age-related differences affected the completion time in all three conditions significantly. However, the measurements specific to cognitive decline and prior knowledge varied in their impact across the three design methods. Prototypes that following image-schematic metaphors alone were less affected by differences in cognitive abilities and prior knowledge compared to the ones designed in the design condition "AD only".

[1]Because the assumption of variance homogeneity was not met for the design conditions, a non-parametric Kruskal-Wallice-test was conducted.

Table 7.2: Quantitative results of study 5 (Mean and [SD]). AD: Affinity Diagram. IS-M: Image-Schematic Metaphors.

Design Method	Age Group	Time [s]	SMEQ	HQS	QUESI
			(0-220)	(1-7)	(1-5)
AD only	younger	174.5 [40.1]	48.38 [35.11]	3.38 [1.01]	3.59 [0.79]
	older	313.3 [33.3]	43.35 [27.93]	4.30 [1.21]	3.47 [0.77]
AD & IS-M	younger	191.3 [51.3]	61.96 [34.37]	4.63 [1.24]	2.78 [0.73]
	older	304.7 [99.2]	45.91 [31.32]	5.02 [1.01]	3.60 [0.59]
IS-M only	younger	150.8 [40.1]	60.35 [36.49]	4.85 [0.96]	2.91 [0.95]
	older	251.8 [75.3]	52.07 [38.71]	5.12 [1.08]	3.18 [1.00]

Table 7.3: Regression models on the influence of age, cognitive abilities and prior technological knowledge on effectiveness and efficiency during the single use cases. Standardised regression coefficients (β) and p-values in brackets. AD: Affinity Diagram. IS-M: Image-Schematic Metaphors. Grey cells: significant amount of variance explained by the predictor.

	AD only		**AD & IS-M**		**IS-M only**	
	Time	SMEQ	Time	SMEQ	Time	SMEQ
R^2	69.3%	6.3%	57.2%	15.9%	54.7%	10.5%
$Adjusted - R^2$	68.2%	3.0%	55.7%	12.8%	53.1%	7.3%
β_{age}	0.42	-.29	.35	-.48	.48	-.34
	*(.001)****	*(.03)**	*(.001)****	*(.001)****	*(.001)****	(.008)^m
$\beta_{cognition}$	-.40	-.20	-.35	-.15	-.27	-.29
	*(.001)****	*(.11)*	*(.001)****	*(.21)*	*(.002)***	*(.02)**
$\beta_{knowledge}$	-.22	-.09	-.26	-.24	-.15	-.13
	*(.002)***	*(.43)*	*(.002)***	*(.03)**	*(.07)*	*(.25)*

Mental effort - SMEQ Over all prototypes, younger adults rated the prototypes as significantly more cognitively demanding than older adults (see Table 7.2), $t(278.5) = 2.38$, $p = .018$, $d = .62$. A comparison of the three design conditions revealed no significantly different scores, $U(2) = 4.7$, $p = .10$. Similar to the statistical analysis of completion times, three regression models were computed to compare the explained variance of age, cognitive abilities and prior knowledge of the perceived mental effort (see Table 7.3. While these variables explained a substantial part of the variance for completion times, they explained less variance for the mental effort ratings. While age explained significant parts of variance for all three design conditions, cognitive abilities was significant only for "IS-M only" and prior knowledge only for "AD & IS-M".

QUESI For the QUESI (measuring the subjectively reported consequences of intuitive use), only 48 data points were collected for younger adults and 46 data points for older adults rendering the calculation of regression models impractical. In an ANOVA with factors "age" and "design condition", no statistically significant effects were found (also not the interaction term). However, the ratings of younger adults were descriptively lower than of older adults. Also, younger adults rated prototypes of the design conditions with image-schematic metaphors as slightly less intuitive, but older adults did not.

However, when comparing the two age groups with each other for each subscale, older adults scored significantly higher on two dimensions: in the subscale "perceived effort of learning", younger adults scored significantly lower ($M = 3.03$, $SD = .94$) than older adults ($M = 3.43$, $SD = .89$), $t(92) = 2.08$, $p = .04$, $d = .44$. Younger adults also scored lower on the dimension of "perceived achievement of goals" ($M = 3.60$, $SD = .96$) compared to older adults ($M = 4.15$, $SD = .69$), $t(92) = 3.16$, $p = .002$, $d = .66$. Younger adults perceived the prototypes as less supportive for the use case compared to older adults. However, when applying alpha-correction following Bonferroni (five subscales leading to a new alpha-level: .01), only the difference for the dimension “perceived achievement” of goals was significant. An ANOVA with revealed no significant differences between design conditions. The five subscales of the QUESI all correlated significantly with each other ($.61 < r < .88$).

7.1.3 Discussion

In study 5, participants evaluated the prototypes from study 3. For this, the paper prototypes of study 3 were further developed into digital prototypes. This development followed an – as much as possible – standardised visual design (based on a stylesheet and clear rules for visual changes) and usability engineering process (to remove severe usability issues, standardise the functionality and increase the comparability of the single prototypes). In a user study, data on perceived innovation, completion time, mental effort and perceived intuitive use were collected for younger and older adults. Also, data on the participants’ cognitive abilities as well as prior technological knowledge were collected. It was expected that prototypes that followed image-schematic metaphors would be a) perceived as more innovative (*H1*) and would decrease the impact of cognitive abilities and prior technological knowledge on user performance (*H3*). Also, it was expected that younger adults would score higher on both cognitive abilities as well as prior technological knowledge measures (*H2*).

As expected, perceived innovation was higher for older adults than for younger adults and higher for prototypes that had been designed with image-schematic metaphors than without. The interaction between these two main effects was not significant. The results support *H1*. The effect of designing with image-schematic metaphors on perceived innovation was larger for younger adults than for older adults. Interestingly, the observation in study 3 that a combination of design methods (affinity diagram and image-schematic metaphors) might have negative consequences for the design process (e.g. lower quality of teamwork) did not lead to less innovative prototypes in this condition.

Younger and older adults differed significantly in their cognitive abilities. However, contrary to the sample of study 1, the two age groups of study 5 did not differ significantly in their self-reported technology competence with the focus on online banking. While younger adults had higher exposure to technology, the self-rated competence was similar compared to older adults, which only partially confirms *H2*. Thus, the effect of integrating image-schematic metaphors in the interfaces (*H3*) might be underestimated in this study.

Regression analysis revealed a significant effect of age on the performance in all conditions. Thus, using image-schematic metaphors in the design process does not lead to general age-inclusiveness. In this study, other age-related factors like sensorimotor differences were left out. Contrary to previous studies (Hurtienne, Klöckner et al., 2015), study 5 also applied measurements that were not based on questionnaires alone but also on objective criteria like completion times. Overall, completion times for prototypes that followed image-schematic metaphors were not significantly dependent on prior technological knowledge in the calculated regression model, but significantly dependent for the other two conditions ("AD only" and "AD & IS-M"). Thus, age-inclusiveness for age-related differences of prior technological knowledge can be partially assumed. Also, the impact of cognitive abilities on the completion time was less for "IS-M only" compared to the other two condition. Still, the impact was significant in all three conditions. In sum, age-inclusiveness based on the objective criterion of completion times was found for the dimension of prior technological knowledge and restrictedly also for cognitive abilities. User interfaces designed with image-schematic metaphors could be assumed to be less dependent on prior technological knowledge.

Younger adults rated the prototypes as significantly more cognitively demanding compared to older adults. This is interesting since the naïve expectation would be the opposite. However, Sonderegger et al. (2016) could show that subjective measurements often do not match objective observation during user testing with older adults, underlining the importance of objective measurements like completion times. Different age groups might interpret the same measurement and given anchors differently (Dickinson et al., 2007; Vines et al., 2015). Thus, one explanation for the lower scores for older adults compared to younger adults regarding cognitive effort could be found in a differing sensitiveness of the SEMQ for younger and older adults. Younger adults were also more familiar with scale-based ratings in user studies because of their background as undergraduate students. Finally, younger adults might have a higher demand for user interfaces in general than older adults based on their prior experiences with modern technologies. The prototypes in this study are not comparable to polished and time-proven user interfaces available today, which might have negatively biased the ratings of younger adults but not of older adults. In sum, the variable of mental effort must be seen critically and cannot be generalised and interpreted in depth in this context. The same argument applies to the questionnaire data stemming from the QUESI where younger adults also scored lower compared to older adults. Several other methodological issues must be noted when interpreting these results.

First, participants did not interact with all prototypes but only with six out of twelve, resulting in an incomplete experimental design. Thus, no dependent tests could be applied. Even though the conducted analyses assume independent data points, they are usually robust also in this case, and analysis of the residuals and residual autocorrelations revealed no violations that would prohibit the application of regression analysis on this dataset.

Second, the prototypes were improved slightly to avoid significant changes in the results of the design process of study 3. An initial expert evaluation of the usability of each prototype led to only minor changes, and further development of the prototypes might affect the results obtained from this study. Also, not all prototypes are optimal and scalable solutions for real-world applications. However, since the aim of this study was to analyse the perceived innovation and the performance for given use cases in a research context, the findings of this study can still be interpreted and serve for further investigations of this topic.

In sum, study 5 provides more empirical data of different age groups while comparing a multitude of interactive prototypes that were developed with and without image-schematic metaphors. Prototypes designed with image-schematic metaphors were perceived as more innovative. Cognitive abilities and prior technological knowledge were shown to be slightly less influential for interacting with image-schematic metaphor prototypes. However, the effects are small.

7.2 Study 6: Comparing IS-M Prototypes with Industry Standards

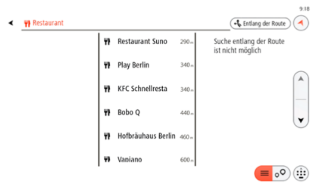
Following the user requirements and image-schematic metaphors obtained from the contextual inquiries of study 2, Winkler et al. (2016) developed prototypes for five use cases. For each of these use cases, they created a) a prototype using image-schematic metaphors and b) a prototype representing the current industry standard. However, their primary user evaluation involved only a small number of participants. Study 6 focused on a more substantial evaluation of these prototypes applying a variety of instruments for screening cognitive abilities and prior technological knowledge of the participants.

7.2.1 Method

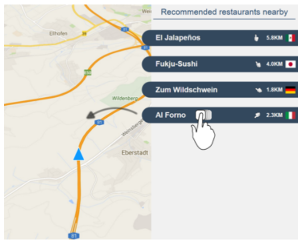
Prototypes Winkler et al. (2016) developed and evaluated eight prototypes (four prototypes based on image-schematic metaphors and four prototypes based on the industry standard, see figure 7.2). They designed half of the prototypes around image-schematic metaphors (e.g., CONVERSATION IS CONTAINER, derived from study 2) and half of the prototypes in accordance to the current industry standard of modern entertainment, communication or navigation systems. These prototypes served as the basis for study 6 with only minor adaptations. For example, an updated version of the used Axure-software avoided technical problems that had occurred in the first study.

The prototypes covered five use cases that were identical between design conditions. In use case 1, participants had to set a favoured restaurant as the new destination along the route. In use case 2, participants had to play a specific song in the playlist. In use case 3, participants had to recommend the song of use case 2 to a friend. In use case 4, participants had to report congestion along the route. In use case 5, participants had to add and remove contacts to and from a group conversation (see also appendix A.5.1 for the instruction for each use case).

Sample 60 participants were recruited for this study (other than in Winkler et al.,

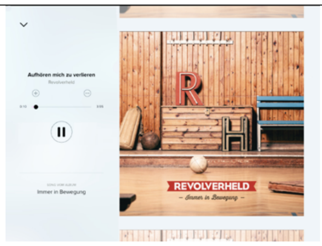


(a) Use Case 1 (Industry Standard)

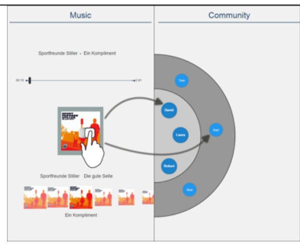


(b) Use Case 1 (IS-M)

ROUTE IS SURFACE



(c) Use Cases 2 and 3 (Industry Standard)



(d) Use Cases 2 and 3 (IS-M)

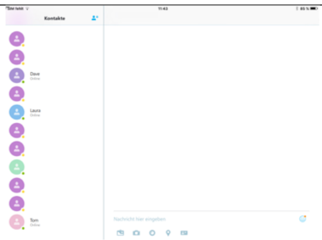
SELECTING SONG IS IN (2)

RECOMMENDATION IS COMPULSION (3)

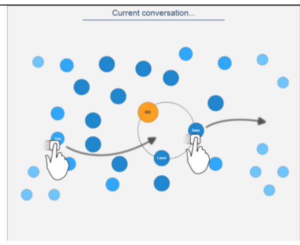
(e) Use Case 4 (Industry Standard)

(f) Use Case 4 (IS-M)

REPORTING CONGESTION IS CONTACT TO PATH



(a) Use Case 5 (Industry Standard)



(b) Use Case 5 (IS-M)

GROUP COMMUNICATION IS CONTAINER

Figure 7.2: Baseline (industry standard) prototypes and IS-M prototypes based on industry standards (Winkler et al., 2016). Note: the white hand was not included in the interactive prototypes and only visualises the user interaction.

2016). The sample included two age-groups. 30 younger adults and 30 older adults. The mean age of younger adults was 22.4 years (*SD* = 2.8; min = 18, max = 28) and 20 of them were female. The mean age of older adults was 55.8 years (*SD* = 4.9; min = 50, max = 68) and 14 of them were female. The education in the group of younger adults was descriptively higher ("Mittlere Reife": 1 participant, "Abitur": 23, "University Degree": 6) than that of older adults ("Hauptschule": 5 participants, "Mittlere Reife": 8, "Abitur": 10, "University Degree": 6).[2] The current occupation differed also between participants. Younger adults were almost exclusively students enrolled at university ("Student": 28 participants, "Soldier": 1, "Engineer": 1). Older adults were mostly still working ("Employee": 14, "Retiree": 4, "Civil Servant/Beamter": 5, "Self-Occupied": 4).[3] Younger adults were recruited via a student panel and received partial course credit or 10 Euro for their participation. Older adults were recruited in local networks or via personal contacts and received 10 Euro for their participation.

Experimental Design Primarily, the study focused on two independent variables. First (design condition), two design conditions were tested: a) prototypes that had been designed explicitly to incorporate single image-schematic metaphors and b) prototypes that followed the industry standard without explicitly incorporating single image-schematic metaphors[4]. Second (age), participants were from the two age groups younger adults and older adults. As described at the beginning of this chapter, two aspects that correlated with age were of particular interest: cognitive abilities (averaged over four markers) and prior technological knowledge (averaged over four markers).

Data for the following dependent variables were collected and analysed (see also section "material"): perceived innovation (HQS-score), performance (completion time, number of clicks) and subjective measures (perceived mental effort via the SMEQ, perceived consequences of intuitive use via the QUESI).

Material For the cognitive assessment, participants completed the same cognitive tests as in study 5: the Trail Making Tests part A and B (Ashendorf et al., 2008; Rabin et al., 2005) as well as the Digit Letter Substitution Test (van der Elst et al., 2006). Again, the three cognitive tests (TMT-A, TMT-B, Digit Letter Substitution Test) were z-standardised and aggregated into one single cognitive abilities score.

Additionally, two more tests were included on an exploratory basis. Reverse Digit Span task and the mini-q. Even though age-differences have not been reported yet for them, these two tests were included for another research group to test whether these measures are also sensitive to age-related differences in cognitive abilities. In the Reverse Digit Span Task (Lezak, 2004), participants were verbally given a sequence of digits (1 – 5 –3) and had to repeat them in a reversed order (3 –5 –1). The number of digits is increased with each iteration (starting with two numbers) until the participant cannot repeat two sequences with the same number of digits. The

[2] The information on the level of education was missing for one older adult.

[3] The information on the current occupation was missing for four older adults.

[4] As we saw in study 4, even designing *without* image-schematic metaphors explicitly can result in the incorporation of few image-schematic metaphors.

	Beispiel			richtig	falsch
1.	Das Dreieck zieht den Kreis vor	■	▲●	☒	☐

Figure 7.3: Example task for the mini-q (translation: "The triangle prefers the circle: right vs wrong.

score is equal to the number of digits the participant was twice able to produce the correct reverse sequence for.

In the mini-q-test (Baudson & Preckel, 2016), participants interpreted triplets of icons (circle, square and triangle) where two of them were closer to each other than the third (see figure 7.3). For each triplet, participants stated whether a sentence describing their alignment was right or wrong (e.g., "the circle is not nearer to the square than the triangle", right or wrong). The final score represents the number of right evaluations for a given time duration (3 minutes). All tests were administered in German.

The assessment of prior knowledge focused on the following four markers that were based on the framework of Hurtienne et al. (2013). Due to time constraints, two levels of competence were not included. The following four markers for prior technological knowledge were z-standardised and aggregated into one single score:

1. Exposure - low specificity. Participants reported for how many years they owned a smartphone.
2. Exposure - medium specificity. Participants reported how often they had performed each of 19 different tasks on a screen-based technology (e.g., listening to music, social media, sending e-mails; "never" to "daily"; range of sum score: 0 to 81).
3. Exposure - high specificity. The familiarity with the industry standard prototypes (e.g. Skype) in this study (range of average score: 1 to 10).
4. Competence - low specificity. Participants completed a questionnaire on technology affinity to electrical devices (TA-EG; Karrer et al., 2009). As in study 1, the subscale "competence" served as a marker for the overall competence with technology (range of average score: 1 to 5).

To measure the perceived innovation of each prototype, the subscale "hedonic quality – stimulation" (HQS) of the AttrakDiff2 was applied (Hassenzahl, 2004). It consists of seven items of a semantic differential including pairs like "innovative – conservative", "novel – conventional" or "bold – cautious". This scale has previously been applied to measure the degree of innovation of a user interface by Hurtienne, Klöckner et al. (2015). Also, the SMEQ was applied to measure perceived mental workload (Sauro & Dumas, 2009) and the QUESI where participants rated the perceived consequences of intuitive use (Naumann & Hurtienne, 2010). Additionally, participants indicated their perceived familiarity with each prototype on a scale ranging from 1 (very unfamiliar) to 10 (very familiar).

Apparatus The interactive prototypes were tested using the Axure-Software on an iPad mini with a diagonal of 7.9inch (20.1cm) and a resolution of 1024x768 pixel (pixel density 163PPI), using the FullScreen Browser Frameless. The luminance of the display (LCD) was set to maximum. The tablet was lying on the table, but participants were also allowed to hold it in their hand for an optimal angle.

Procedure Each session of the study was conducted individually at the university's laboratory or, to increase comfort for older adults, at the participant's home. After collecting demographic data and prior technological knowledge, the tests TMT-A and -B, Digit Span Reversed, Digit Letter Substitution and mini-q were administered (see section "material"). Insights from similar studies including cognitive assessment led to the recommendation of applying cognitive ability tests before the main experiment (Chevalier et al., 2012). Also, the experimenter introduced all participants to the tablet-based interaction style (e.g. swipe gesture, drag and drop) and explained more details when more information was necessary. This introduction also served as a short test technical test of whether participants could read text on the tablet easily and whether the tablet recognised their touch gestures.[5]

First, participants completed all five use cases with the first design condition. The instruction (see also appendix A.5.1) for the use cases was identical to Winkler et al. (2016) and did not differ between design conditions. The design condition of the first block of prototypes was randomised. After each use case, participants rated a) the perceived innovation (HQS), b) their perceived mental effort (SMEQ) and c) their perceived familiarity of the prototype. After the first block of five prototypes with one design condition, participants completed the QUESI questionnaire and repeated the same procedure for the second design condition. Finally, participants were given the screens of the two corresponding prototypes for each use case and chose, which of the two they would favour in the future.

7.2.2 Results

Data analysis was conducted using SPSS version 24 and Excel 2016. Alpha-level was set to .05 and differences in degrees of freedom occurred, for example, due to the violation of assumptions like sphericity or homogeneity of variances.

Perceived innovation Figure 7.4 summarises the perceived innovation over all prototypes and age groups. Over all trials, the reported familiarity ratings significantly correlated negatively with the HQS-score, $r(598) = -.35$, $p < .001$. To analyse the effect of age (younger vs. older adults), design condition (image-schematic metaphors vs. industry standard) and use case (five), a 2 x 2 x 5 mixed ANOVA was conduc-

[5] This is often a physical problem of older adults interacting with touchscreens (Barnard et al., 2013).

ted. The main effect of age was significant and younger adults in general reported a higher perceived innovation than older adults, $F(1{,}58) = 12.30$, $p = .001$, $n_p^2 = .18$. Also, the main effect of design condition was significant and prototypes conforming with image-schematic metaphors were perceived as more innovative than prototypes based on the industry standard, $F(1{,}58) = 47.87$, $p < .001$, $n_p^2 = .45$. Finally, the main effect of use case was significant and participants perceived the use cases as differently innovative, $F(4{,}232) = 12.79$, $p < .001$, $n_p^2 = .18$.

Additionally, all second-order interactions were significant. The impact of design condition was larger for younger adults than for older adults, $F(1{,}58) = 23.28$, $p < .001$, $n_p^2 = .29$, the impact of design condition varied over the use cases, $F(4{,}232) = 6.23$, $p < .001$, $n_p^2 = .10$, and the impact of age varied between different use cases, $F(4{,}232) = 5.30$, $p = .001$, $n_p^2 = .08$. The three-fold interaction between age, design condition and use case was not significant, $F(4{,}232) = 1.44$, $p = .23$.

Table 7.4: Perceived innovation per use case and design condition for both age groups (Mean and [SD]).

Use Case	Age Group	Younger Adults	Older Adults
1	Standard	3.40 [0.67]	4.77 [0.92]
	IS-M	4.59 [0.66]	4.70 [0.75]
2	Standard	4.17 [0.98]	4.53 [1.04]
	IS-M	5.09 [0.81]	4.88 [1.01]
3	Standard	3.52 [1.10]	4.63 [1.16]
	IS-M	5.29 [0.83]	5.20 [0.94]
4	Standard	4.27 [0.73]	5.11 [0.80]
	IS-M	5.09 [0.68]	5.09 [0.89]
5	Standard	3.84 [0.98]	5.07 [0.83]
	IS-M	5.15 [0.82]	5.30 [0.73]

Cognitive abilities As reported in literature, younger adults performed better in

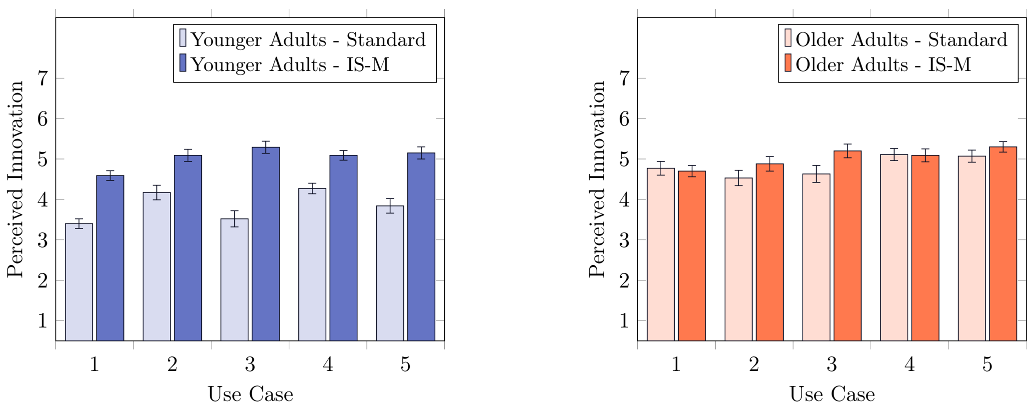


Figure 7.4: Mean and Standard Error of perceived innovation per age group, design condition and use case. IS-M: Image-schematic metaphor prototype. See also table 7.4.

the three tests TMT-A, TMT-B and Digit Letter Substitution Test compared to the age group of older adults. In the TMT-A, younger adults were significantly faster in completing the test (M = 24.7s, SD = 6.9s) compared to older adults (M = 34.5s, SD = 10.9s), $t(49.2) = 4.14$, $p < .001$, $d = 1.07$ (degrees of freedom adjusted due to variance inhomogeneity). Also the TMT-B, younger adults were significantly faster in completing the test (M = 51.7s, SD = 17.9s) compared to older adults (M = 78.9s, SD = 32.7s), $t(44.9) = 4.00$, $p < .001$, $d = 1.03$ (degrees of freedom adjusted due to variance inhomogeneity). In the Digit Letter Substitution Test, younger adults substituted significantly more letters with digits ($M = 42.6$, $SD = 7.3$) compared to older adults ($M = 33.4$, $SD = 4.9$), $t(58) = 5.73$, $p < .001$, $d = 1.48$. The overall z-standardized cognitive score (younger adults: $M = .113$, $SD = .366$; older adults: $M = -.113$, $SD = .498$) ranged from -.76 to +1.33 for this sample.

Younger and older adults did not differ significantly in the other two measures mini-q (younger adults: $M = 29.4$, $SD = 13.5$; older adults: $M = 31.7$, $SD = 11.9$; $t(57) = .71$, $p = .48$) and Digit Span Reversed (younger adults: $M = 8.3$, $SD = 2.2$; older adults: $M = 7.4$, $SD = 2.1$; $t(58) = 1.6$, $p = .11$). Because they had been included only on an explorative basis for another research team and no age-differences have been reported in literature, they were not included in the overall cognitive score.[6]

Prior technological knowledge Younger adults reported significantly higher levels of technology competence - low specificity (e.g. TA-EG questionnaire; $M = 3.74$, $SD = .72$) than older adults ($M = 2.98$, $SD = .81$), $t(58) = 3.87$, $p < .001$, d=.99. Also on the dimension of exposure - low specificity, younger adults scored higher (usage of a smartphone; $M = 3.97$, $SD = 2.99$) compared to older adults ($M = 1.80$, $SD = 1.94$), $t(58) = 3.35$, $p = .001$, $d = .86$. Contrary to the expectations, the score for exposure - medium specificity (e.g., frequency of performing a set of tasks with technology), was significantly lower for younger adults ($M = 27.8$, $SD = 13.8$) compared to older adults ($M = 40.6$, $SD = 21.0$), $t(50.0) = 2.79$, $p = .007$, $d = .72$. Finally, younger adults scored higher for exposure - high specificity (familiarity with the industry standard prototypes, $M = 7.14$, $SD = 1.66$) than older adults ($M = 3.43$, $SD = 2.24$), $t(56) = 7.17$, $p < .001$, $d = 1.89$. Based on the four markers, an overall prior knowledge-score was calculated which was z-standardised (younger adults: $M = .31$, $SD = .47$; older adults: $M = -.28$, $SD = .65$). For this sample the score ranged from -1.38 to +1.19.

Performance To analyse the influence of the three predictors age, cognitive abilities and prior technological knowledge, several regression models were computed. For the binary criterion "use case completion" (yes/no), a binary logistic regression model was used, while linear regression models were used for the interval scaled criteria "completion time", "subjective mental effort" and "perceived consequences of intuitive use" (see table 7.5).

Use case completion. Overall, age and age-related differences could explain only a

[6]However, the assumption that these tests would serve as an indicator for age-related changes in cognitive abilities must be discarded.

small proportion of the variance in the objective performance measures. For each of the two design conditions, a binary logistic regression with use case completion (yes/no) as the binary criterion and the cognition score, prior knowledge score and age as the three predictors was computed. The overall explained variance of the models was low (Standard: 14.8%; IS-M: 11.9%), and for both design conditions, age was a significant predictor of the probability of use case completion. However, while the cognition score (*cognition* in table 7.3) and the prior knowledge score (*knowledge* in table 7.5) explained a significant amount of variance in the logistic regression for the industry standards, this was not the case for the logistic regression model for the IS-M prototypes.

Completion time. As in the regression model on use case completion, the overall explained variance of the models was low (Standard: 11.5%; IS-M: 9.0%). In the model of industry standard prototypes, the cognition score and prior knowledge score had a significant influence, but not age alone. In the model of IS-M prototypes, age and the cognition score had a significant influence, but not the prior knowledge score.

Mental effort. Two linear regression models with age, cognition score and prior knowledge score were also calculated for the subjectively perceived mental effort, operationalised via the SMEQ-scale. The overall explained variance was also low for this variable (Standard: 11.4%; IS-M: 13.7%). Here, both age and prior knowledge explained significant variance in both models, but not the cognition score.

Perceived consequences of intuitive use. Finally, two linear regression models were computed for the QUESI-scores. Compared to the other measurements, the predictors explained more variance (Standard: 33.7%; IS-M: 12.9%). While the factors age and the prior knowledge score were significant in the model for standard prototypes, none of the three factors was significant in the model of IS-M prototypes.

7.2.3 Discussion

Study 6 focused on the comparison of interactive prototypes that were based on the image-schematic metaphors from study 2 with industry standards as a baseline. Younger and older adults were screened regarding their cognitive abilities and prior technological knowledge and completed standardised use cases. It was expected that participants would perceive prototypes designed following image-schematic metaphors as more innovative compared to industry standard prototypes. Also, cognitive abilities and prior technological knowledge should influence the participant's performance for image-schematic metaphor prototypes to a lesser degree than for the industry standard prototypes.

In fact, participants rated image-schematic metaphor prototypes as more innovative, which supports *H1*. This effect mainly stems from higher innovation ratings of younger adults. Older adults did not differ in their innovation ratings of image-schematic metaphor and industry standard prototypes. This finding might be explained by the

Table 7.5: Regression models on the influence of age, cognitive abilities and prior technological knowledge on effectiveness and efficiency during the single use cases. Use case completion: logistic regression with standardised coefficients calculated after Menard (2011). Other variables: linear regression. Standard: Prototypes based on the industry standard. IS-M: Prototypes based on specific image-schematic metaphors. Brackets represent p-values. Grey cells indicate predictors that explain a significant amount of variance.

	Completion (y/n)		Time	SMEQ	QUESI
Standard					
$Nagelkerke-R^2$	14.8%	R^2	11.5%	11.4%	33.7%
$log-likelihood$	285.69	$AdjustedR^2$	10.7%	10.5%	30.1%
β_{age}	-0.16	β_{age}	0.12	0.19	-0.23
	*(.02)**		*(.06)^m*	*(.003)***	*(.08)^m*
$\beta_{cognition}$	-0.10	$\beta_{cognition}$	0.17	-0.07	-0.02
	(.08)^m		*(.005)***	*(.22)*	*(.88)*
$\beta_{knowledge}$	0.14	$\beta_{knowledge}$	-0.16	-0.22	0.43
	*(.03)**		*(.001)***	*(.001)***	*(.002)***
IS-M					
$Nagelkerke-R^2$	11.9%	R^2	9.0%	13.7%	12.9%
$log-likelihood$	310.1	$AdjustedR^2$	8.1%	12.8%	8.0%
β_{age}	-0.75	β_{age}	0.17	0.21	-0.2
	*(.001)****		*(.008)***	*(.001)****	*(.18)*
$\beta_{cognition}$	-0.20	$\beta_{cognition}$	0.18	-0.07	-0.04
	(.27)		*(.003)***	*(.24)*	*(.20)*
$\beta_{knowledge}$	0.10	$\beta_{knowledge}$	-0.04	-0.24	0.2
	(.62)		*(.52)*	*(.001)****	*(.20)*

tablet-based interaction which was often novel to older adults compared to younger adults. Still, following image-schematic metaphors as a basis for interaction design led to non-standard interface elements, a more visual form of data visualisation and gesture-based user input.

Younger and older adults of this sample differed significantly in their cognitive abilities and prior technological knowledge, which supports also *H2*. The higher variance in cognitive abilities in the group of older adults compared to younger adults is in line with the literature (Ashendorf et al., 2008). Because incorporating image-schematic metaphors in the interface allows the participant to exploit technology-independent prior knowledge with a minimum of cognitive resources, the hypothesis was that differences in participants' cognitive abilities, and prior technological knowledge would less influence and thus explain less variance in image-schematic metaphor prototypes than in the industry standard prototypes. This effect was found in the data, but to a smaller extent than expected.

Age had a significant effect in all but one regression models, which was expected because this variable also includes sensorimotor differences between age groups. Contrary to cognitive abilities and prior technological knowledge, the factor age subsumed a variety of different age-related variables that were not the focus of this study and thus not explicitly measured (e.g. motor speed, perceptual speed).

Interaction design with image-schematic metaphors was expected to reduce the influence of differences in cognitive abilities. For use case completion rates, cognitive abilities was a marginally significant factor in the industry standard prototypes, but not in the image-schematic metaphor prototypes. This at least partially supports *H3*, even though, for completion times, cognitive abilities were significant for both types of prototypes, and no influence of cognitive abilities was found for perceived mental effort and the perceived consequences of intuitive use.

More importantly, prior technological knowledge was a significant predictor in the regression models of all four criteria (use case completion, completion time, mental effort, intuitive use) for the industry standard, but only for one for the IS-M prototypes (mental effort). This is promising since it indicates that the interaction (effectiveness and efficiency) with the image-schematic metaphors prototypes is less dependent on prior technological knowledge and thus decreases the impact of one core age-difference (*H3*).

Taken together, this study showed two things for prototypes that followed image-schematic metaphors: on the one hand, they were perceived as more innovative by younger adults (but not by older adults) leading to age-independent high innovation ratings. Moreover, on the other hand, the amount of variance that was explained by prior technological knowledge might be reduced due to interaction design with image-schematic metaphors. However, several methodological limitations and alternative explanations to these findings have to be considered when interpreting these results.

First, the same interaction design team developed all prototypes, making the process prone to biases. This is undoubtedly a weakness of this study. However, this approach allowed for a systematic comparison that was not affected by differences between design teams and itself already provides an improvement compared to previous work without baselines (Antle et al., 2009; Hurtienne, Klöckner et al., 2015).

Second, using industry standards (e.g. Skype or navigation systems available in the market) as a baseline also had the consequence that participants with more prior technological knowledge could apply their knowledge to the industry standard prototypes but not the image-schematic metaphor prototypes. This could explain the missing impact of prior technological knowledge scores on image-schematic metaphor prototypes. Still, this study aimed to show that image-schematic metaphors provide a tool for designing interfaces that are invariant to age-related differences. This aim must be regarded as fulfilled.

Third, the study sample included older adults that some authors would not label as older adults (A. Smith, 2014). The resulting small age-difference might underestimate the differences between younger and older adults in this study. Even though the focus of this study was on cognitive abilities and prior technological knowledge, which have been shown to decline already earlier than 50 years (Czaja et al., 2006;

Salthouse, 2010), also “older” older adults should be addressed in the future.

7.3 Summary

This chapter focused on two main promises of image-schematic metaphors in interaction design: innovation and age-inclusiveness. In two user studies, participants evaluated interactive prototypes that were either explicitly based on image-schematic metaphors or not. Table 7.6 summarises the outcome of this chapter.

Both studies support the first hypothesis (*H1*) stating that participants should perceive image-schematic metaphor prototypes as more innovative. This finding is in line with Hurtienne, Klöckner et al. (2015) who also showed in a study without a baseline that their developed prototype (based on image-schematic metaphors) was perceived as innovative. However, in the studies described in this chapter, older adults perceived already the baselines as innovative and, therefore, the effect of increased perceived innovation mainly stems from higher innovation ratings of younger adults.

The second hypothesis (*H2*) stated that younger and older adults would differ in their cognitive abilities and report lower prior technological knowledge. This hypothesis was also confirmed for both samples of the two studies, and younger and older adults significantly differed in these two dimensions. However, the differences in prior technological knowledge were small.

H2 was not new and age-related differences starting early in the lifespan have been well documented in literature (Baltes & Lindenberger, 1997; Reddy et al., 2014; Salthouse, 2010). However, directly measuring these age-related differences in the samples were essential for the third hypothesis (*H3*) stating that age-related differences such as cognitive resources and prior technological knowledge should influence HCI less in image-schematic metaphor prototypes than in baseline prototypes. The results indicate that IS-M prototypes are not age-inclusive in general and age was a significant predictor in all regression models. However, this work focused on two particular dimensions of age-inclusiveness: cognitive abilities and prior technological knowledge. In both studies, the impact of differences in prior technological knowledge but not in cognitive abilities was decreased by incorporating image-schematic metaphors in both studies. Taken together, the promise of image-schematic metaphors to provide age-inclusiveness (Hurtienne, Klöckner et al., 2015) is met primarily for the inter-individual differences of prior technological knowledge.

The two studies are the first reported in the literature that systematically compared image-schematic metaphor prototypes to baselines in a standardised experimental setting with a large number of participants and different prototypes. Thus, the primary contribution of this chapter is providing empirical data for two core promises of image-schematic metaphor theory that have been neglected until now. Demographic change emphasises the importance of user interfaces that can be interacted with

Table 7.6: Overview of the results regarding the hypotheses of studies 5 and 6. IS-M: image-schematic metaphors.

	Hypothesis	Study 5	Study 6
H1	Higher perceived innovation of IS-M prototypes than baseline	✓	✓
H2	Sample with age-differences regarding:		
	- Cognitive abilities	✓	✓
	- Prior technological knowledge	✓/χ	✓
H3	Higher age-inclusiveness of IS-M prototypes than baseline regarding:		
	- Cognitive abilities	χ	✓/χ
	- Prior technological knowledge	✓	✓

independently of prior technological knowledge and cognitive resources. Also, designing with image-schematic metaphors could provide a tool for achieving both this age-independent interaction while increasing its perceived innovation, which has been shown to be a primary driver in product design (Radford & Bloch, 2011; E. M. Rogers, 1995). Despite several methodological constraints that limit the generalisability of these results, this work is still a strong indicator that image-schematic metaphors can be successfully applied in interaction design to achieve innovative and – regarding differences in prior technological knowledge – age-inclusive HCI at the same time.

Chapter 8

Conclusion

It is time to connect the findings of the studies of this thesis and return to the big picture. Image-schematic metaphors, in theory, promise to foster both innovative as well as age-inclusive interaction design. As outlined in chapter 4, following image-schematic metaphors in the design process facilitates ideas on the border between tradition (knowledge about the world) and transcendence (conventional mechanisms presented in unconventional ways, Biskjaer et al., 2010); image-schematic metaphors help to converge or distill the users' mental models into actionable elements while sparking divergent thinking during the ideation process; they provide structure in the design process as a meta-language for interaction design, and, finally, they act as sources of inspiration (Hurtienne, Klöckner et al., 2015). In terms of age-inclusive design, image-schematic metaphors capture technology-independent prior knowledge that is accessible to humans of different ages and with differences in cognitive abilities and prior technological knowledge. Basing designs on such metaphors should increase the likelihood that various users can interact with a system intuitively. The studies of this work were designed to test both promises empirically.

Besides aggregating the core findings of this work, this chapter aims at discussing strengths and weaknesses of this work on a higher level. As a conclusion, the synthesis of the method of image-schematic metaphors in interaction design from chapter 4 is developed into a short manual that presents the core steps that are needed for its application in practice. Finally, possible directions for future work are discussed.

8.1 Scope: the aim of this work

The focus of this thesis was on testing two core promises of the method of image-schematic metaphors: integrating image-schematic metaphors in interaction design processes should lead to both innovative as well as age-inclusive user interfaces. It is already challenging to develop user interfaces that are perceived as innovative. Additionally, these innovative user interfaces may disadvantage especially older adults, although the demographic change makes the consideration of older age groups a necessity. Examples such as the use of an unnatural set of affordances to interact with a system (e.g., in mid-air gestures) or decreasing screen sizes can contradict accessibility guidelines, which is a reason why innovativeness and age-inclusiveness are often investigated independently or result into entirely different systems. However, a user interface designed to be age-inclusive can also be potentially perceived as innovative, which is why the combination of both goals is a core contribution of this work.

The thesis aimed to apply the approach of image-schematic metaphors – originally stemming from cognitive linguistics – to this two-fold problem. Contrary to age-inclusion in general (including sensorimotoric differences), the focus of this work was on differences between younger and older adults regarding cognitive abilities and prior technological experience. As described in chapter 2, these two factors are essential for HCI. And, as described in chapter 4, incorporating image-schematic metaphors in interaction design can potentially lead to innovative and – concerning cognitive abilities and prior technological knowledge – age-inclusive interfaces (Hurtienne, Klöckner et al., 2015; Winkler et al., 2016). However, the literature provides three vital research gaps:

First, the argumentation of age-inclusive interfaces relied primarily on the assumption that younger and older adults make use of the same metaphors in their language. Until now, this was an untested assumption not yet supported by empirical data (*R1*).

Second, as the method of image-schematic metaphors can not yet be considered as widespread or industry standard, designers have to invest time in learning and adapting it into their portfolio. Moreover, the method itself is competing against a lot of other more established design approaches that aim at facilitating innovation or age-inclusiveness. In order to reach wide dissemination, the method needs to add value not only to the design outcome, but also during the design process itself. Previous work often did not go beyond qualitative feedback from single project teams, sometimes even without a baseline design condition, which made interpretation of the obtained results difficult. Therefore, the value of following image-schematic metaphors was compared to standard design processes with parameters characterising the design process (e.g. perceived applicability, perceived creativity) and the outcome (e.g. perceived innovation of the designed user interfaces, number of generated design ideas)(*R2*).

Third, empirical evaluations of interactive prototypes that had been designed with or without image-schematic metaphors were limited or non-existent. By this, previous findings of innovative user interfaces created with image-schematic metaphors were difficult to interpret (e.g. Hurtienne, Klöckner et al., 2015). Additionally, the impact of participants' differences in cognitive abilities and prior technological experience and its interaction with the design process of the prototypes were not subject to investigation before (*R3*).

This work addressed all three research questions to provide insights into innovative and age-inclusive interaction design with image-schematic metaphors. While research questions one and three primarily address theoretical concerns, all three research questions aimed at providing insights that can also be useful for practitioners. By explicitly focusing on single steps of the extraction, design and evaluation process, larger samples and a comparison of multiple prototypes per design condition than in previous studies were made possible.

The main findings of this thesis make both empirical and methodological contributions. Firstly, the empirical data provides preliminary answers to yet untested theoretical assumptions. These findings are necessary for the empirical foundation of the method of image-schematic metaphors for interaction design in general, and, more specifically, for innovative and age-inclusive interaction design. Secondly, the empirical findings also have implications of methodological nature. Insights from applying the method of image-schematic metaphors allow to draw conclusions on how to improve the approach in practice and guide future applications of the method of image-schematic metaphors.

8.2 Empirical Contributions

8.2.1 Age-universality of image-schematic metaphors

Chapter 5 addressed the question of whether and to which extent image-schematic metaphors are universal across different ages (*R1*). The prediction was that – due to the experiential foundation of image schemas and image-schematic metaphors (sensomotoric level) – younger and older adults would show the same image-schematic metaphors in their natural language, even though they should differ in their prior technological experience (expertise level).

By applying a highly standardised procedure, study 1 revealed a substantial overlap of image-schematic metaphors in the natural language of younger and older adults. In total, the overlap was 75%, even though the two age groups differed in their prior technological experience (in this case online banking). Following a more applied methodology of collecting natural language in the domain of social communication between vehicles, study 2 could replicate this overlap (70%).

These findings largely support the claim from previous work (Hurtienne, Klöckner et al., 2015) that knowledge from lower levels of the knowledge continuum – based on sensorimotoric knowledge – is largely shared among different user groups compared to more specific expert knowledge from higher levels (Hurtienne & Blessing, 2007). However, the overlap is still far from being perfect. Around 3 of 10 image-schematic metaphors were not shared between younger and older adults. Most of these metaphors were unique to one age group (mostly the younger adults) and in only few cases the same target domain was instantiated with different image-schematic metaphors.

This means that the mismatch between the found image-schematic metaphors was less because of inconsistencies, but rather due to the more metaphor-rich language of younger adults. This finding is in line with previous research, showing that the overlap between different languages in bi-lingual individuals is substantial, but not perfect, and mainly due to unique metaphors instead of competing ones (Löffler et al., 2014). This mismatch in image-schematic metaphors can be seen as additional source for design, rather than a thread to age-inclusive design when following metaphors that are only present in one age group. Thus, from a practical stance, the underlying motivation of grounding user interfaces in image-schematic metaphors in the users language is fulfilled.

8.2.2 Designers' perceived value of integrating image-schematic metaphors in the design process

Chapter 6 focused on the new method's applicability and potential as an source of inspiration from the designers' perspective (*R2*). In study 3 and 4, novices and experts in interaction design applied the method of image-schematic metaphors in the design process under different conditions. Study 3 for the first time allows for a comparison of the new method with the industry standard Contextual Design (more specifically, the phase of affinity diagram) (Holtzblatt & Beyer, 2017). Study 4 extended the target group to professional interaction designers and was, contrary to previous work (Hurtienne, 2011; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Maier et al., 2013), not constrained to novices. This allows for an estimation of the expected acceptance of image-schematic metaphors in projects working with professional interaction designers. Based on the results, novice and expert interaction designers are likely to perceive image-schematic metaphors as a readily applicable method for focusing on the user's mental models while at the same time boosting creativity and out-of-the-box thinking: while the standard method of affinity diagramming only led to about 22 ideas in study 3, the image-schematic metaphors condition resulted in 70 design ideas. Additionally, professional interaction designers did not deviate to a large extent from their standard procedure (observed in the baseline team of study 4), but were able to integrate image-schematic metaphors in the form of an assistance method for the standard procedure.

Most of the previous HCI-related research involving image-schematic metaphors is positioned on the gap between software requirements and interface design and did not focus on innovative solutions that are also age-inclusive. Studies 3 and 4, therefore, provide the comparison of different design processes and expand existing recommendations for designing with image-schematic metaphors with insights gained from a larger sample size of designers. Taken together, these insights are promising with regard to the acceptance of the method in practice.

8.2.3 Evaluating image-schematic metaphor prototypes for innovation and age-inclusiveness

In design research, new methods for interaction design are often not tested against baselines rendering statements of its advantages compared to other methods challenging. Designing with image-schematic metaphors promises two core advantages over existing approaches: innovative and age-inclusiveness regarding cognitive abilities and prior technological knowledge (*R3*). Studies 5 and 6 provide empirical data on participants interacting with a multitude of prototypes that were designed with or without image-schematic metaphors. Combined with an extensive assessment of the participants' cognitive abilities and prior technological knowledge, these user studies allowed for investigating these two age-related differences between younger and older adults.

Study 5 revealed that prototypes were less affected by differences in cognitive abilities and prior technological knowledge when they had been designed with image-schematic metaphors. Furthermore, these prototypes were also perceived as more innovative by younger and older adults. This supports the claim of Hurtienne (2017a) that image-schematic metaphors in interaction design might lead to age-inclusive and at the same time innovative interfaces. Also in study 6, the hypotheses were confirmed, and prototypes that were based on image-schematic metaphors were perceived as more innovative, and performance was less dependent on cognitive abilities and prior technological knowledge. Studies 5 and 6 show that designing with image-schematic metaphors in fact are a appropriate approach to innovative and age-inclusive interaction design Still,, the contribution of the method to age-inclusiveness seems to be limited to differences in prior technological knowledge and not cognitive ageing.

8.3 Methodological Contributions

The second contribution of this thesis lies in gathering insights from applying the method of image-schematic metaphors in practice. The following sections will first summarise the critical methodological takeaways from the studies of this thesis (section 8.3) and afterwards merge these new insights with current recommendations that have been proposed by previous applications of image-schematic metaphors in practice (section 8.5).

Interaction design can benefit from integrating image-schematic metaphors into the design process. For example, as one outcome of study 4, existing design procedures should not be overruled but enriched by image-schematic metaphors. The final goal is to increase the practicality of finding domain-specific image-schematic metaphors and blending them into the design process. The first set of recommendations supports and facilitates the finding age-inclusive image-schematic metaphors. The second set of recommendations encourages and facilitates the integration of image-schematic metaphors in the design process.

Finding age-inclusive image-schematic metaphors. The first recommendation for the extraction process focuses on the question whether only younger adults are sufficient for extracting age-inclusive image-schematic metaphors to guide the interaction design for abstract concepts. As we saw in studies 1 and 2, younger and older adults show a substantial overlap in their use of image-schematic metaphors. For example, in study 1, younger and older adults used the same image-schematic metaphors most frequently to talk about the same abstract concept in 75% of the abstract concepts. This means that the probability of using the most frequent image-schematic metaphor found for an abstract concept in the group of younger adults is likely to be age-inclusive (see figure 5.5 in chapter 5). However, the chance might be that exactly this image-schematic metaphor is not age-inclusive. For example, younger and older adults used different image-schematic metaphors to talk about friendship (LINKAGE vs ENABLEMENT). To minimise the chance of erroneously designing under the wrong assumption of an age-inclusive image-schematic metaphor, material from both younger and older adults should be included in the extraction phase.

The second recommendation addresses the composition of the final set of image-schematic metaphors that should serve as the basis for interaction design. Study 2 revealed that image-schematic metaphors are more likely to be age-inclusive when it is used by more than one person of an age-group. This is partially in line with the recommendation of Löffler, Hess, Hurtienne et al. (2013), who argue that only image-schematic metaphors with multiple occurrences qualify for being used in the interaction design. However, image-schematic metaphors with multiple occurrences in *different people* should be prioritised over image-schematic metaphors with numerous occurrences in *the same person*. Based on the results of study 2, finding an image-schematic metaphor in transcripts of two younger people sets the probability of also finding it in older adults to about 50% and finding an image-schematic metaphor in three younger adults increases the likelihood to 100%. Even though these numbers should not be generalised without restrictions, this is an important takeaway that amends one of the core recommendations of Löffler, Hess, Hurtienne et al. (2013).

The third recommendation concerns the instruction of the team that extracts metaphors from the transcripts. For a given abstract concept, sufficient inter-rater reliability for extracting image-schematic metaphors highly unstructured material of study 2 was found (Krippendorf's $\alpha = .70$). This is interesting since the raters differed greatly in their extraction style. While one coder analysed all syllables and words that apparently were of no interest for the project and interaction design, the two other coders were more efficient in their extraction. Thus, it might be worth first to define the abstract concepts that are important to the project before starting the extraction process. This recommendation is in line with previous work (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013). However, due to the specific research question, it was not possible to follow this recommendation in study 2. Study 1, on the other hand, was geared towards a highly efficient extraction methodology. Due to the standardised questions to the participants, the abstract concept was already set, and the extractors did not have to define the abstract concept and extract the image schemas. Instead, extracting image schemas directly led to image-schematic metaphors.

Taken together, the phase of extracting image-schematic metaphors from natural language can follow one of two paths, which could be labelled as “quick and practical” and “thorough and more valid”. When the project-relevant abstract concepts are already evident, the approach of study 1 might be useful. Asking a variety of participants to describe the concepts of interest produces mostly linguistic material that is relevant for precisely these concepts of interest. In combination with the short length of the answers makes extraction very efficient. However, this approach might not reveal all image-schematic metaphors for the concept but only those that are most available to the participant. Also, in practice, it might be difficult to get together a high number of participants as in study 1. However, this approach is, as long as the abstract concepts can be defined before collecting the linguistic material and participants are available, quick and practical, but narrow and thus possibly invalid with a smaller number of participants. In comparison, the approach of study 2, which uses the transcripts of classic contextual interviews, can also be applied when the project-relevant abstract concepts are not clear yet. Additionally, it will more likely provide a more diverse linguistic material with fewer participants. The approach of extracting image-schematic metaphors from transcripts is, therefore, more thorough and valid but requires more resources in the extraction process. To make this approach more feasible, one should focus only on design-relevant material or user utterances relevant for core functionality.

Integrating image-schematic metaphors in the design process. The first recommendation for integrating image-schematic metaphors in the design process focuses on its combination with other design methodologies. The image-schematic metaphors are seen as a useful method and participants in study 3 and 4 found ways of working with them. Most importantly, in study 3, participants did not rate the method of image-schematic metaphors as less applicable than a standard design methodology like Affinity Diagramming. However, a Wall Walk with image-schematic metaphors led to better teamwork in study 3 than when the image-schematic metaphors were blended into the Wall Walk of an Affinity Diagram. Since interaction designers often follow their own or the company’s procedures (see study 4), the standard method should remain the primary focus. Image-schematic metaphors can only enhance and enrich the standard procedure in practice to provide the interaction designers with an additional source of inspiration and make explicit age-inclusive mental models of later users.

The second recommendation concerns the strictness of the method, more explicitly the force of integrating all given image-schematic metaphors into the interface. Interaction designers in study 4 ignored one of the provided image-schematic metaphors and incorporated two very late. In study 3, participants stated that not all metaphors could be easily integrated with the interface. This means two things: On the one hand, the list of metaphors can be concise so that the entire interface can benefit from every single one. On the other side, during the instruction of the method, it might be highlighted that integrating all metaphors is not mandatory. This is already what professional interaction designers did naturally in study 4, but which novices, who sticked to the standardised instructions in the laboratory setting, found challenging about the method.

The third recommendation concerns whether to conduct a wall walk with image-schematic metaphors or not. The wall walk with image-schematic metaphors resembles the wall walk during contextual design and to do it requires time. In study 3, participants remarked that they came up with many useless design ideas during the wall walk with image-schematic metaphors. Thus, practitioners might be tended to skip the wall walk and just present the image-schematic metaphors for inspiration. For two reasons, the wall walk should be included when possible. First, the brainstorming of many ideas contains useless design ideas in most interaction design frameworks and is an accepted part of interaction design (Dark Horse Innovation, 2016). Removing apparently irrelevant design ideas from the process might remove precisely those that stimulated the most fruitful ones in later stages of the design process. Second, study 3 showed that image-schematic metaphors are prone to be dominated by an affinity diagram. Only including them in the Wall Walk of the affinity diagram undermines the potential of image-schematic metaphors. Instead, a separate Wall Walk with only image-schematic metaphors might be a better approach.

8.4 Recap of the Research Methodology

Several limitations, which have been already discussed in the single studies (see chapters 5 to 7), have to be taken into account when interpreting these results and recommendations. This section aims at summarising the most critical ones and discussing research decisions on a more fundamental level. By this, the section tries to make explicit the reasons for the chosen methodologies of the presented studies and emphasise their respective advantages and disadvantages.

As outlined in the introduction (see chapter 1), the methodology adopted in this thesis tries to balance on different dimensions to investigate the set research questions. Taken together, five major points for discussion are relevant from a high-level view: “theory vs practice”, “experimental standardisation (internal validity) vs generalisability (external validity)”, “depth (one prototype tested and analysed in detail) vs breadth (many prototypes from different conditions)”, “quantitative vs qualitative” and “novices vs professionals”.

Theory vs practice Besides describing the theoretical predictions of image-schematic metaphors for innovative and age-inclusive interaction design in chapter 4, the thesis provides empirical data for testing one of the core assumptions of previous work: that younger and older adults overlap in their use of image-schematic metaphors in language. However, showing that empirical data meet this theoretical assumption is not directly relevant for interaction designers. More importantly, the thesis shows *that* and especially *how* interaction design with image-schematic metaphors can lead to more innovative and age-inclusive user interfaces compared to standard design approaches.

Experimental standardisation vs generalisability Especially when the subject of research is interaction design, the balance between experimental standardisation (internal validity) and generalisation (external validity) must be carefully considered (Zimmerman & Forlizzi, 2014). Thus, each research question was addressed by two studies: one that aimed at a high standardisation but with arguable limitations of external validity from the perspective of an interaction designer (studies 1, 3 and 6) and one that aimed at a high potential for generalisation but with arguable limitations from the perspective of an experimental researcher (studies 2, 4 and 5). This approach addressed both perspectives, but also served for replicating findings by using different research methods.

Few vs many prototypes Previous research focused mostly on only one or very few image-schematic metaphor prototypes (Hurtienne, 2011; Hurtienne, Klöckner et al., 2015). Moreover, evaluations against baselines were limited (Asikhia et al., 2015; Hurtienne, 2011; Löffler, Hess, Maier et al., 2013). The thesis adopted the approach of comparing multiple image-schematic metaphor prototypes against multiple prototypes that were not explicitly based on image-schematic metaphors. Even though this approach entailed its own limitations (e.g. a less standardised experimental manipulation), the increased number of prototypes (in total 20 in studies 5 and 6) was expected to make the results less affected by random artefacts of baseline choice.

Quantitative vs qualitative HCI-research usually involves quantitative measurements while design research often relies on qualitative data and explorative analysis, which is difficult to connect to the empirical sciences (Zimmerman et al., 2007). The studies reported in this thesis apply, where possible, quantitative measurements, but include, where useful, qualitative information. For example, in studies 3 and 4, qualitative data provided a deeper insight into the design process, insights quantitative measures could not have captured. Qualitative approaches also allowed to get insights into details of the extraction and design process on an exploratory basis. Thus, the major findings of this study are of quantifiable nature, but qualitative insights enrich the statements of this thesis.

Novices vs professionals Finally, the research described in this thesis sought to include participants that are difficult to recruit from conventional student panels. More specifically, the studies required older adults as well as professional interaction designers. Students were included as novice interaction designers (with a background in degree courses in interaction design or human-computer interaction) and younger adults because they were easier to recruit, which allows for larger sample sizes. However, they might have lacked the professional experience that is relevant for applying the method of image-schematic metaphors in practice. Also, the younger adults included in this research were affine to technology, which might have affected the comparison with older adults. Professional interaction designers (and older adults) are more challenging to recruit and allow for only small sample sizes.

Even though it would be favourable to avoid all possible limitations, the research approach sometimes has to prioritise one goal over another. The six studies presented in this thesis try to cover different poles to provide insights into the topic of innovative and age-inclusive interaction design with image-schematic metaphors. Each pole of the dimensions comes with its unique advantages and disadvantages, but to further develop the picture drawn by previous work, all of these dimensions had to be regarded. Also, these dimensions are tightly linked to each other. However, the multi-method approach of this thesis was necessary for addressing the set research questions and progressing both empirical foundations for theoretical assumptions as well as the practicality of the method of image-schematic metaphors.

8.5 Synthesis revisited: Designing with image-schematic metaphors

Finally, it is time to wrap up the methodology of designing with image-schematic metaphors for innovative and age-inclusive interaction design. This section aims to interweave the findings of this thesis with previous recommendations and provide a comprehensive summary of the overall process for researchers and practitioners who want to apply the method of image-schematic metaphors in projects. A detailed step-by-step manual for practitioners can be found in appendix A.6. The following section makes accessible the contribution of this thesis in combination with yet undocumented best practices from previous work (e.g. unpublished additional material from projects; Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013) and helps to promote and further develop the methodology, materials and procedures.

As we saw in chapter 4 (see section 4.4), interaction design with image-schematic metaphors involves primarily two phases: an extraction phase and an ideation/prototyping phase. More specifically, the project team must first identify relevant metaphors. Afterwards, the project team integrates this list of metaphors into the particular user interface.

8.5.1 Extraction of image-schematic metaphors

To find image-schematic metaphors, chapter 4 discussed several sources. When possible, existing image-schematic metaphors from existing databases should be exploited

(Hurtienne, 2017b). It is important to emphasise that the method of image-schematic metaphors in its essence promises to support interaction design for mapping abstract concepts to physical dimensions and processes. When no domain-specific metaphors are documented, the project team must carry out language analysis itself. For this, they can use one of two methods:

Standardised: Descriptions of abstract concepts. When the core concepts are already pre-defined (e.g. in the project scope) and no user research has been done yet, potential users can be instructed to describe these concepts in their own words. This minimises the effort necessary for transcription and metaphor analysis, but there is the chance that fruitful and possibly relevant image-schematic metaphors will be omitted. This method was introduced in study 1.

User-Centered Design: Contextual interviews. When the core concepts are not pre-defined, or the project team wants to find even more image-schematic metaphors, the transcripts of contextual interviews, conducted during user research, are a good source to start with. Importantly, the extraction of image-schematic metaphors should only start after the core concepts have crystalised. Parts of the transcripts that do not cover these core concepts should not be coded. However, this approach still requires more resources than the standardised approach because all contextual interviews must be transcribed completely to find exactly these locations that contain expressions on the relevant concepts. This method was introduced earlier (Hurtienne, Klöckner et al., 2015) and also adopted in study 2.

While the standardised approach was introduced in study 1, the contextual interviews approach was applied in most projects until now (Hurtienne, Klöckner et al., 2015; Löffler, Hess, Hurtienne et al., 2013). In line with previous work, the contextual interviews approach was resource-consuming. A combination of these two approaches might be optimal, which is asking participants during usual contextual interviews to describe abstract concepts that are likely to become relevant to the project. By transcribing and analysing only these descriptions at first, the project team has fast and efficient access to image-schematic metaphors. If additional concepts emerge during the project that has not been part of the participants' descriptions, the project team can – when resources are available – transcribe and analyse the contextual interviews to find also image-schematic metaphors for these additional concepts.

To increase the probability that image-schematic metaphors are age-universal, natural language of both younger and older adults should be elicited in the extraction phase. In total, the number of participants should follow the usual sample sizes of User-Centered Design, which is between 4 and 12 participants (Holtzblatt & Beyer, 2017). When only the standardised descriptions are used as a basis for the extraction process, the number should be higher (e.g. at least ten younger and ten older adults) to ensure a rich basis of natural language.

8.5.2 Ideation and prototyping

The image-schematic metaphors extracted in the first phase have to be translated into concrete design solutions. Two aspects must be considered carefully in this context. First, the method of image-schematic metaphors should stand alone for itself. The impact of image-schematic metaphors is decreased when they are introduced in existing design procedures like affinity diagramming. Thus, designing with image-schematic metaphors should be detached from other procedures to facilitate the adoption of the method. The brainstorming of instantiations of image-schematic metaphors in the user interface should be a separate activity, which allows the team members to focus on the representation of abstract concepts in the interface that conform with the metaphors. The metaphors are comparable to inspiration cards that stimulate the creativity during the design process. However, when these sources of inspiration are neglected due to other distracting methods of interaction design, the potential of image-schematic metaphors is not fully exploited.

Second, the expertise of the team members that are responsible for interaction design should be taken into account. Professional interaction designers might be capable of integrating image-schematic metaphors more easily in existing design procedures. They will ignore single image-schematic metaphors when the resulting interface would admittedly conform with it, but become too odd leading to poor user experience and usability. Novices that can draw on only a little or no experience with interaction design might stick to the image-schematic metaphors and try to integrate all given ones into the interface leading to sometimes odd interfaces that violate basic heuristics for good usability. Thus, for novices, the set of image-schematic metaphors should be smaller and more focused than for professionals.

In sum, several projects integrated image-schematic metaphors into the design process, but the accessibility to the insights from these projects to practitioners and research is limited. Appendix A.6, therefore, provides a brief step-by-step manual that explains the concrete steps that are relevant for applying the method of image-schematic metaphors in practice. In combination with existing materials (Löffler, Hess, Hurtienne et al., 2013), this should also allow novices to integrate image-schematic metaphors in their projects to stimulate creativity. As we saw in chapter 7, grounding user interfaces in image-schematic metaphors by applying the reported procedures can increase both the perceived innovation of the user interface as well as decrease the impact of age-related differences when users interact with these user interfaces.

8.6 Conclusion and Outlook

Image-schematic metaphors promise to facilitate both innovative as well as age-inclusive interaction design. Previous work could not provide sufficient empirical data that this promise might be fulfilled. Two studies in this thesis (study 1 and 2) showed that image-schematic metaphors are universal across different age groups. Following the theory that language reveals underlying mental representations, this means that interfaces that instantiate these image-schematic metaphors directly tap into universal prior knowledge supporting HCI. By applying the method of image-schematic metaphors, two more studies (studies 3 and 4) gathered feedback from a large group of participants and compared the new method to a baseline. This feedback is a valuable source for promoting and improving the method of image-schematic metaphors in practice. The recommendations derived from these two studies were interwoven with insights from previous projects that integrated image-schematic metaphors in the interaction design process. The distilled step-by-step manual can support practitioners and researchers that are unfamiliar with the method but have no access to experts in the method. The final two studies (studies 5 and 6) positively answered the question whether prototypes that were designed with image-schematic metaphors were, in fact, more innovative and more age-inclusive than baselines. Interaction with prototypes that were designed with image-schematic metaphors was less dependent on prior technological knowledge and cognitive abilities, two crucial dimensions age-related differences between users typically affect.

However, these studies still provide room for improvement and extensions. Even though the basis is laid for innovative and age-inclusive interaction design using image-schematic metaphors, additional questions have arisen that call for more research in different directions.

First, the thesis focused on two age groups that do not represent the complete span of age – from children to oldest older adults. Future work could consider samples from older age groups than in this thesis. This is necessary to replicate the findings of, on the one hand, age-universality of image-schematic metaphors in language and, on the other hand, the advantages of image-schematic metaphors in age-inclusive interaction design.

Second, future work should also focus on a micro-level and how precisely image-schematic metaphors operate in the user interface. For example, to which extend the amount of explicitely designed image-schematic metaphors affects age-inclusiveness and perceived innovation, or how image-schematic metaphors differ from primary metaphors, which are even more deeply rooted at the sensorimotor stage of prior knowledge. This work provides the framework for applying the method in practice and sheds light on the general process of designing for innovative and age-inclusive interaction design. However, work on the impact of single image-schematic metaphors on specific interface elements is still rare, but essential for further strengthening the foundation of interaction design with image-schematic metaphors.

Third, more research needs to be undertaken regarding the efficiency of the extraction of image-schematic metaphors from natural language. Building on previous work, this thesis provides a standard procedure that aims at supporting the application of the method in projects. However, the intention of providing such a standard procedure is also to trigger discussions about the concrete implementation. Following the philosophy of iterative User-Centered Design, deriving lessons learned from real applications and field-tests is important for making progress. Thus, the proposed procedure is not intended to be final, but instead a basis for future improvements of the methodology.

Fourth and finally, the method of designing for innovation and with image-schematic metaphors needs further attention regarding its advantages and disadvantages compared to other design methods. Even though experimental comparisons to other design methods are usually rare in design research, more design methods than User-Centred Design are available in practice and literature. Future work could compare the method of image-schematic metaphors with other approaches to innovative and age-inclusive interaction design to replicate the findings of this work.

To conclude, this work investigated core assumptions and promises of interaction design with image-schematic metaphors. This thesis fills yet open research gaps, but also extends the method to practice by analysing natural language from diverse age groups, including diverse groups in the interaction design process, applying diverse methodologies in the studies and comparing various prototypes to baselines. In its essence, this work intends to move the method of image-schematic metaphors closer to the application in practice by building on theory.

Bibliography

Ackerman, P. L. & Rolfhus, E. L. (1999). The locus of adult intelligence: Knowledge, abilities, and nonability traits. *Psychology and Aging*, *14*(2), 314–330.

Adams, R., Bessant, J. & Phelps, R. (2006). Innovation management measurement: A review. *International Journal of Management Reviews*, *8*(1), 21–47.

Antle, A. N., Corness, G. & Droumeva, M. (2009). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interacting with Computers*, *21*(1-2), 66–75.

Arbuckle, T. Y. & Gold, D. P. (1993). Aging, inhibition, and verbosity. *Journal of Gerontology*, *48*(5), 225–232.

Ashendorf, L., Jefferson, A. L., O'Connor, M. K., Chaisson, C., Green, R. C. & Stern, R. A. (2008). Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology*, *23*(2), 129–137.

Asikhia, O. K. & Setchi, R. (2016). Linking Design Intention and Users' Interpretation through Image Schemas. *IFAC-PapersOnLine*, *49*(19), 283–288.

Asikhia, O. K., Setchi, R., Hicks, Y. & Walters, A. (2015). Conceptual framework for evaluating intuitive interaction based on image schemas. *Interacting with Computers*, *27*(3), 287–310.

Aula, A., Khan, R. M. & Guan, Z. (2010). How does search behavior change as search becomes more difficult? In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 35–44). New York: ACM.

Bäckman, L., Small, B. J. & Wahlin, A. (2001). Aging and memory. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 349–377). San Diego, CA: Academic Press.

Baddeley, A. (1986). Modularity, mass-action and memory. Special Issue: Human memory. *Quarterly Journal of Experimental Psychology Human Experimental Psychology*, *38*(4), 527–533.

Baddeley, A. (2001). Is working memory still working? *American Psychologist*, *56*(11), 851–864.

Baddeley, A. & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89.

Baddeley, A. & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mech-*

anisms of active maintenance and executive control. (pp. 28–61). New York: Cambridge University Press.

Bailey, C. & Sheehan, C. (2009). Technology, older persons' perspectives and the anthropological ethnographic lens. *Alter*, *3*(2), 96–109.

Bakker, S., Antle, A. N. & van den Hoven, E. (2009). Identifying embodied metaphors in children's sound-action mappings. In *Proceedings of the 8th international conference on interaction design and children - idc '09* (pp. 140–149).

Bakker, S., Antle, A. N. & Van Den Hoven, E. (2012). Embodied metaphors in tangible interaction design. *Personal and Ubiquitous Computing*, *16*(4), 433–449.

Baltes, P. B. & Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: a new window to the study of cognitive aging? *Psychology and aging*, *12*(1), 12–21.

Baltes, P. B., Mayer, K. U., Helmchen, H. & Steinhagen-Thiessen, E. (2010). Überblick und Einführung. In U. Lindenberger, J. Smith, K. U. Mayer & P. B. Baltes (Eds.), *Die berliner altersstudie* (pp. 25–58). Berlin: Akademie Verlag.

Baregheh, A., Rowley, J. & Sambrook, S. (2009). Towards a multidisciplinary definition of innovation. *Management Decision*, *47*(8), 1323–1339.

Barnard, Y., Bradley, M. D., Hodgson, F. & Lloyd, A. D. (2013). Learning to use new technologies by older adults: Perceived difficulties, experimentation behaviour and usability. *Computers in Human Behavior*, *29*(4), 1715–1724.

Barnett, H. G. (1953). *Innovation: the basis of cultural change* (1st ed.). New York: McGraw-Hill.

Baudson, T. G. & Preckel, F. (2016). Mini-q: A three-minute intelligence screening. *Diagnostica*, *62*(3), 182–197.

Bertram, L., Böckenhoff, A., Demuth, I., Düzel, S., Eckardt, R., Li, S. C., ... Steinhagen-Thiessen, E. (2014). Cohort profile: The Berlin aging study II (BASE-II). *International Journal of Epidemiology*, *43*(3), 703–712.

Birren, J. E., Woods, A. M. & Williams, M. V. (1980). Behavioral slowing with age: Causes, organization, and consequences. In L. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 293–308). Washington, DC, US: American Psychological Association.

Biskjaer, M. M., Dalsgaard, P. & Halskov, K. (2010). Creativity Methods in Interaction Design. In *Proceedings of the 1st desire network conference on creativity and innovation in design* (pp. 16–17). Lancaster, UK: DESIRE Network.

Biskjaer, M. M., Dalsgaard, P. & Halskov, K. (2017). Understanding Creativity Methods in Design. In *Proceedings of the 2017 conference on designing interactive systems* (pp. 839–851). New York: ACM.

Biskjaer, M. M. & Halskov, K. (2014). Decisive constraints as a creative resource in interaction design. *Digital Creativity*, *25*(1), 27–61.

Blackler, A. & Hurtienne, J. (2007). Towards a unified view of intuitive interaction: definitions, models and tools across the world. *MMI-interaktiv*, *13*, 36–54.

Blackler, A., Mahar, D. & Popovic, V. (2010). Older adults, interface experience and cognitive decline. In *Proceedings of the 22nd conference of the computer-human interaction special interest group of australia on computer-human interaction - ozchi '10* (pp. 172–175). New York: ACM.

Blackler, A. & Popovic, V. (2015). Editorial: Towards intuitive interaction theory. *Interacting with Computers*, *27*(3), 203–209.

Blackler, A., Popovic, V., Mahar, D. P., Reddy, R. & Lawry, S. (2012). Intuitive interaction and older people. In P. Israsena, J. Tangsantikul & D. Durling (Eds.), *Proceedings of the design research society (drs) 2012 conference* (pp. 560–578). Bangkok: Chulalongkorn University.

Blackwell, A. F. (2006). The reification of metaphor as a design tool. *Transactions on Computer-Human Interaction*, *13*(4), 490–530.

Boot, W. R., Charness, N., Czaja, S. J., Sharit, J., Rogers, W. A., Fisk, A. D., ... Nair, S. (2013). Computer proficiency questionnaire: assessing low and high computer proficient seniors. *The Gerontologist*, *55*(3), 404–411.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F. & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, *44*(2), 123–147.

Botwinick. (1977). Intellectual abilities. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 580–605). New York: Van Nostrand Reinhold.

Brajnik, G., Yesilada, Y. & Harper, S. (2011). Web accessibility guideline aggregation for older users and its validation. *Universal Access in the Information Society*, *10*(4), 403–423.

Brébion, G. (2001). Language processing, slowing, and speed/accuracy trade-off in the elderly. *Experimental Aging Research*, *27*(2), 137–150.

Brebion, G., Smith, M. J. & Ehrlich, M.-F. (1997). Working memory and aging: deficit or strategy differences? *Aging, Neuropsychology, and Cognition*, *4*(1), 58–73.

Bruder, C., Blessing, L. & Wandke, H. (2014). Adaptive training interfaces for less-experienced, elderly users of electronic devices. *Behaviour & Information Technology*, *33*(1), 4–15.

Burke, D. & Shafto, M. (2008). Language and Aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *Handbook of aging and cognition* (pp. 373–443). New Jersey: Lawrence Erlbaum Associates, Inc.

Burke, K. (1962). *A Grammar of Motives and a Rhetoric of Motives.* Cleveland: World Publishing Company.

Caldwell, B., Cooper, M., Reid, L. G. & Vanderheiden, G. (2008). *Web Content Accessibility Guidelines (WCAG) 2.0* (Tech. Rep.). Retrieved from `http://www.w3.org/TR/WCAG20/`

Carmien, S. & Manzanares, A. G. (2014). Elders using smartphones - A set of research based heuristic guidelines for designers. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8514 LNCS*(PART 2), 26–37.

Casasanto, D. & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of educational psychology*, *54*(1), 1.

Center, P. R. (2015). *The Smartphone Difference.* Retrieved from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Cerella, J. (1985). Information processing rates in the elderly. *Psychological bulletin*, *98*(1), 67–83.

Chattopadhyay, D. & Bolchini, D. (2015). Motor-intuitive interactions based on image schemas: Aligning touchless interaction primitives with human sensorimotor abilities. *Interacting with Computers*, *27*(3), 327–343.

Chayko, M. (2014). Techno-social Life: The Internet, Digital Technology, and Social Connectedness. *Sociology Compass*, *8*(7), 976–991.

Chen, K. & Chan, A. H. S. (2013). Use or non-use of gerontechnology-A qualitative study. *International Journal of Environmental Research and Public Health*, *10*(10), 4645–4666.

Chen, K. & Chan, A. H. S. (2014). Gerontechnology acceptance by elderly Hong Kong Chinese: a senior technology acceptance model (STAM). *Ergonomics*, *57*(5), 635–652.

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

Chevalier, A., Dommes, A. & Marquié, J.-C. (2015). Strategy and accuracy during information search on the Web: Effects of age and complexity of the search questions. *Computers in Human Behavior*, *53*, 305–315.

Chevalier, A., Dommes, A. & Martins, D. (2012). The effects of ageing and website ergonomic quality on internet information searching. *Ageing and Society*, *33*, 1009–1035.

Chulvi, V., Mulet, E., Chakrabarti, A., López-Mesa, B. & González-Cruz, C. (2012). Comparison of the degree of creativity in the design outcomes using different design methods. *Journal of Engineering Design*, *23*(4), 241–269.

Coleman, R., Clarkson, P. J., Dong, H. & Cassim, J. (2016). *Design for inclusivity: a practical guide to accessible, innovative and user-centred design.* Boca Raton, Florida, USA: CRC Press.

Connelly, S. L., Hasher, L. & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging*, *6*(4), 533–541.

Cooper, A. (1995). The myth of metaphor. *Visual Basic Programmer's Journal*, *5*(6), 127–128.

Copeland, D. E., Bies-Hernandez, N. J. & Gunawan, K. (2016). The role of cognition on age-related changes in language, memory, and mental models. *Cognition, Language and Aging*, 135.

Coyne, R. (2005). Wicked problems revisited. *Design Studies*, *26*(1), 5–17.

Crabb, M. (2013). Human cognitive measurement as a metric within usability studies. In *Chi'13 extended abstracts on human factors in computing systems* (pp.

2677–2682). New York: ACM.

Crabb, M. & Hanson, V. L. (2014). Age, technology usage, and cognitive characteristics in relation to perceived disorientation and reported website ease of use. In *Proceedings of the 16th international acm sigaccess conference on computers & accessibility* (pp. 193–200). New York: ACM.

Craik, F. I. M. & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional response. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York: Plenum Press.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Craik, F. I. M. & Salthouse, T. A. (2008). *The handbook of aging and cognition*. New York: Psychology Press.

Crilly, N., Moultrie, J. & Clarkson, P. J. (2004). Seeing things: consumer response to the visual domain in product design. *Design studies*, *25*(6), 547–577.

Csikszentmihalyi, M. (1997). *Flow and the psychology of discovery and invention*. New York: Harper Perennial.

Czaja, S. J., Charness, N., Fisk, A. D., Hertzog, C., Nair, S. N., Rogers, W. A. & Sharit, J. (2006). Factors predicting the use of technology: Findings from the center for research and education on aging and technology enhancement (create). *Psychology and Aging*, *21*(2), 333–352.

Czaja, S. J. & Lee, C. (2007). The impact of aging on access to technology. *Universal Access in the Information Society*, *5*(4), 341–349.

Czaja, S. J. & Sharit, J. (1998). Age differences in attitudes toward computers. *The journals of gerontology. Series B, Psychological sciences and social sciences*, *53*(5), 329–340.

Dahlin, K. B. & Behrens, D. M. (2005). When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy*, *34*(5), 717–737.

Dalsgaard, P. (2014). Design Thinking and Innovation in HCI. In *Proceedings of the 8th nordic conference on human-computer interaction* (pp. 5–7). New York: ACM.

Damodaran, L., Olphert, C. W. & Sandhu, J. (2014). Falling off the bandwagon? Exploring the challenges to sustained digital engagement by older people. *Gerontology*, *60*(2), 163–173.

Dark Horse Innovation. (2016). *Digital Innovation Playbook*. Berlin: Murmann Publishers.

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, *13*(3), 319–340.

Dee, M. & Hanson, V. L. (2014). A large user pool for accessibility research with representative users. In *Proceedings of the 16th international acm sigaccess conference on computers & accessibility* (pp. 35–42). New York: ACM.

Deignan, A. (2006). *Metaphor and Corpus Linguistics*. Amsterdam: John Benjamins Publishing Company.

Dickinson, A., Arnott, J. & Prior, S. (2007). Methods for human-computer interaction research with older people. *Behaviour & Information Technology*, *26*(4), 343–352.

Diefenbach, S. & Ullrich, D. (2015). An experience perspective on intuitive interaction: Central components and the special effect of domain transfer distance. *Interacting with Computers*, *27*(3), 210–234.

Dijkstra, K. & Misirlisoy, M. (2009). Recognition accuracy for original and altered verbal memory reports in older adults. *The Quarterly Journal of Experimental Psychology*, *62*(2), 248–256.

Dinar, M., Summers, J. D., Shah, J. & Park, Y.-S. (2016). Evaluation of Empirical Design Studies and Metrics. In P. Cash, T. Stankovic & M. Storga (Eds.), *Experimental design research* (pp. 13–39). Berlin: Springer.

Djajadiningrat, J., Gaver, W. & Fres, J. (2000). Interaction relabelling and extreme characters: methods for exploring aesthetic interactions. In *Proceedings of the designing interactive systems conference (dis)* (pp. 66–71). New York: ACM.

Docampo Rama, M., Ridder, H. D. & Bouma, H. (2001). Technology generation and age in using layered user interfaces. *Gerontechnology*, *1*(1), 25–40.

Dodge, E. & Lakoff, G. (2005). Image schemas : From linguistic analysis to neural grounding. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 57–91). Berlin: Mouton de Gruyter.

Doucet, L. & Jehn, K. A. (1997). Analyzing harsh words in a sensitive setting: American expatriates in communist China. *Journal of Organizational Behavior*, *18*(S1), 559–582.

Duden. (2017). *Duden online - die deutsche Rechtschreibung.* Retrieved from `www.duden.de/`

Dunne, D. & Martin, R. (2006). Design thinking and how it will change management education: An interview and discussion. *Management Learning and Education*, *5*(4), 512–523.

Eckert, C. & Stacey, M. (2000). Sources of inspiration: a language of design. *Design Studies*, *21*(5), 523–538.

Ehn, P. (1988). *Work-oriented design of computer artifacts (Doctoral Dissertation).* Umea, Sweden: Almqvist and Wiksell International for Arbetslivcentrum.

Elliot, A. J. & Maier, M. a. (2014). Color psychology: effects of perceiving color on psychological functioning in humans. *Annual review of psychology*, *65*, 95–120.

Erickson, T. D. (1995). Working with interface metaphors. In *The art of human-computer interface design* (pp. 65–73). San Francisco, CA: Morgan Kaufmann Publishers Inc.

Fauconnier, G. & Turner, M. (2008). *The way we think: Conceptual blending and the mind's hidden complexities.* Basic Books.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175–191.

Fisk, A. D., Rogers, W. A., Charness, N., Czaja, S. J. & Sharit, J. (2009). *Designing for older adults: Principles and creative human factors approaches* (1st ed.). Boca Raton, Florida, USA: CRC Press.

Förster, J., Marguc, J. & Gillebaart, M. (2010). Novelty categorization theory. *Social and Personality Psychology Compass*, *4*(9), 736–755.

Fozard, J. L., Gordon-Salant, S., Birren, J. E. & Schaie, K. W. (2001). Changes in vision and hearing with aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging (5th ed.).* (pp. 241–266). San Diego, CA: Academic Press.

Freelon, D. (2013). ReCal OIR: Ordinal, interval, and ratio intercoder reliability as a web service. *International Journal of Internet Science*, *8*(1), 10–16.

Fry, A. F. & Hale, S. (2000). Relationships among processing speed, working memory, and fluid intelligence in children. *Biological psychology*, *54*(1-3), 1–34.

Gephart, R. P. (1993). The Textual Approach: Risk and Blame In Disaster Sensemaking. *Academy of Management Journal*, *36*(6), 1465–1514.

Gibson, C. B. & Zellmer-Bruhn, M. E. (2001). Metaphors and Meaning: An Intercultural Analysis of the Concept of Teamwork. *Administrative Science Quarterly*, *46*(2), 274–303.

Goggin, N. L. & Stelmach, G. E. (1990). Age-Related Deficits in Cognitive-Motor Skills. *Advances in Psychology*, *72*(C), 135–155.

Goldschmidt, G. & Rodgers, P. A. (2013). The design thinking approaches of three different groups of designers based on self-reports. *Design Studies*, *34*(4), 454–471.

Goldsmith, R. E. & Hofacker, C. F. (1991). Measuring consumer innovativeness. *Journal of the Academy of Marketing Science*, *19*(3), 209–221.

Gooberman-Hill, R. & Ebrahim, S. (2007). Making decisions about simple interventions: Older people's use of walking aids. *Age and Ageing*, *36*(5), 569–573.

Grégoire, J. & Van der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology and Cognition*, *4*(2), 140–149.

Gromann, D. & Hedblom, M. M. (2017). Kinesthetic Mind Reader: A Method to Identify Image Schemas in Natural Language. In *Proceedings of advances in cognitive systems* (pp. 5–14).

Guilford, J. (1967). *The nature of human intelligence* (1st ed.). New York: McGraw-Hill.

Haegerstrom-Portnoy, G., Schneck, M. E., Lott, L. A., Hewlett, S. E. & Brabyn, J. A. (2014). Longitudinal Increase in Anisometropia in Older Adults. *Optometry and Vision Science*, *91*(1), 60–67.

Hanington, B. & Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions* (1st ed.). Beverly, MA: Rockport Publishers.

Hanson, V. L. (2011). Technology skill and age: What will be the same 20 years from now? *Universal Access in the Information Society*, *10*(4), 443–452.

Hardill, I. & Olphert, C. W. (2012). Staying connected: Exploring mobile phone use amongst older adults in the UK. *Geoforum*, *43*(6), 1306–1312.

Hargittai, E. & Dobransky, K. (2017). Old Dogs, New Clicks: Digital Inequality in Skills and Uses among Older Adults. *Canadian Journal of Communication*, *42*(2), 196–212.

Hart, S. G. (2006, 10). Nasa-Task Load Index (NASA-TLX) 20 Years Later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, pp. 904–908). SAGE PublicationsSage CA: Los Angeles, CA.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, *52*(C), 139–183.

Hasan, M. S. & Yu, H. (2017). Innovative developments in HCI and future trends. *International Journal of Automation and Computing*, *14*(1), 10–20.

Hasher, L., Lustig, C. & Zacks, R. (2008). Inhibitory Mechanisms and the Control of Attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). Oxford: Oxford University Press.

Hasher, L., Stoltzfus, E. R., Zacks, R. T. & Rypma, B. (1991). Age and inhibition. *Journal of experimental psychology. Learning, memory, and cognition*, *17*(1), 163–169.

Hasher, L. & Zacks, R. T. (1988). Working Memory, Comprehension, and Aging: A Review and a New View. *Psychology of Learning and Motivation - Advances in Research and Theory*, *22*(C), 193–225.

Hassenzahl, M. (2003). The Thing and I: Understanding The Relationship Between User and Product. In M. Blythe, K. Overbeeke, A. Monk & P. Wright (Eds.), *Funology: from usability to enjoyment* (pp. 31–42). Berlin: Springer.

Hassenzahl, M. (2004). The Interplay of Beauty, Goodness, and Usability in Interactive Products. *Human-Computer Interaction*, *19*(4), 319–349.

Hawthorn, D. (2000). Possible implications of aging for interface designers. *Interacting with Computers*, *12*(5), 507–528.

Hawthorn, D. (2007). Interface design and engagement with older people. *Behaviour & Information Technology*, *26*(4), 333–341.

Hayes, A. F. & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89.

Hekkert, P., Snelders, D. & Wieringen, P. C. W. (2003). 'Most advanced, yet acceptable': Typicality and novelty as joint predictors of aesthetic preference in industrial design. *British journal of Psychology*, *94*(1), 111–124.

Herstatt, C., Kohlbacher, F. & Bauer, P. (2011). "Silver" product design: Product innovation for older people. *Working Papers / Technologie- und Innovationsmanagement, Technische Universitaet Hamburg-Harburg*, *65*, 2–14.

Hirschman, E. C. (1980). Innovativeness, novelty seeking, and consumer creativity. *Journal of Consumer Research*, *7*(3), 283–295.

Hollins, B. & Pugh, S. (1990). *Successful product design: what to do and when.*

London: Butterworth-Heinemann.

Holtzblatt, K. & Beyer, H. (2017). *Contextual design : design for life* (2nd ed.). Burlington, Massachusetts, USA: Morgan Kaufmann Publishers Inc.

Holtzblatt, K., Wendell, J. B. & Wood, S. (2005). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Burlington, Massachusetts, USA: Morgan Kaufmann.

Hope, A., Schwaba, T. & Piper, A. M. (2014). Understanding digital and material social communications for older adults. In *Proceedings of the 32nd annual acm conference on human factors in computing systems - chi '14* (pp. 3903–3912).

Howard, J. & Howard, D. (1997). Learning and memory. In W. A. Rogers & A. D. Fisk (Eds.), *Handbook of human factors and the older adult* (pp. 7–26). San Diego, CA: Academic Press.

Hughes, L. D., Done, J. & Young, A. (2011). Not 2 old 2 TXT: There is potential to use email and SMS text message healthcare reminders for rheumatology patients up to 65 years old. *Health Informatics Journal*, *17*(4), 266–276.

Hurtienne, J. (2011). *Image Schemas and Design for Intuitive Use (Doctoral Dissertation)*. Berlin: TU Berlin.

Hurtienne, J. (2017a). How Cognitive Linguistics Inspires HCI: Image Schemas and Image-Schematic Metaphors. *International Journal of Human–Computer Interaction*, *33*(1), 1–20.

Hurtienne, J. (2017b). *ISCAT*. Retrieved from `http://iscat.psyergo.uni-wuerzburg.de`

Hurtienne, J. & Blessing, L. (2007). Design for Intuitive Use - Testing Image. In J.-C. Bocquet (Ed.), *Proceedings of the international conference on engineering design iced'07* (pp. 1–12). Paris: Design Society.

Hurtienne, J., Horn, A. M., Langdon, P. M. & Clarkson, P. J. (2013). Facets of prior experience and the effectiveness of inclusive design. *Universal Access in the Information Society*, *12*(3), 297–308.

Hurtienne, J., Israel, J. H. & Weber, K. (2008). Cooking up real world business applications combining physicality, digitality, and image schemas. In *Proceedings of the 2nd international conference on tangible and embedded interaction* (p. 239). New York: ACM.

Hurtienne, J., Klöckner, K., Diefenbach, S., Nass, C. & Maier, A. (2015). Designing with image schemas: Resolving the tension between innovation, inclusion and intuitive use. *Interacting with Computers*, *27*(3), 235–255.

Hurtienne, J. & Langdon, P. M. (2010). Keeping warm in winter: Image-schematic metaphors and their role in the design of central heating controls. In *Proceedings of the fourth international conference of the german cognitive linguistics association* (pp. 53–54). Bremen: University of Bremen.

Hurtienne, J., Löffler, D., Gadegast, P. & Hußlein, S. (2015). Comparing Pictorial and Tangible Notations of Force Image Schemas. In *Proceedings of the ninth international conference on tangible, embedded, and embodied interaction - tei '14* (pp. 249–256). New York: ACM.

Hurtienne, J., Stößel, C., Sturm, C., Maus, A., Rötting, M., Langdon, P. M. & Clarkson, P. J. (2010). Physical gestures for abstract concepts: Inclusive design with primary metaphors. *Interacting with Computers*, *22*(6), 475–484.

Im, S. & Workman, J. P. (2004). Market orientation, creativity, and new product performance in high-technology firms. *Journal of marketing*, *68*(2), 114–132.

Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G. & Petersen, R. C. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. *The Clinical Neuropsychologist*, *10*(3), 262–278.

Jacob, R. J., Girouard, A., Hirshfield, L. M., Horn, M. S., Shaer, O., Solovey, E. T. & Zigelbaum, J. (2008). Reality-based interaction. In *Proceeding of the twenty-sixth annual chi conference on human factors in computing systems - chi '08* (p. 201).

Jayroe, T. J. & Wolfram, D. (2012). Internet searching, tablet technology and older adults. In *Proceedings of the asist annual meeting* (Vol. 49, pp. 1–3).

Jetter, H.-C., Reiterer, H. & Geyer, F. (2014). Blended Interaction: understanding natural human—computer interaction in post-WIMP interactive spaces. *Personal and Ubiquitous Computing*, *18*(5), 1139–1158.

Johannessen, J.-A., Olsen, B. & Lumpkin, G. (2001). Innovation as newness: what is new, how new, and new to whom? *European Journal of Innovation Management*, *4*(1), 20–31.

Johansson-Sköldberg, U., Woodilla, J. & Çetinkaya, M. (2013). Design thinking: Past, present and possible futures. *Creativity and Innovation Management*, *22*(2), 121–146.

Johnson, J. & Finn, K. (2017). *Designing User Interfaces for an Aging Population: Towards Universal Design* (1st ed.). Burlington, Massachusetts, USA: Morgan Kaufmann.

Johnson, M. (1987). *The Body in the mind: The bodily basis of meaning, reason and imagination.* Chicago, Illinois, USA: University of Chicago Press.

Kanagal, N. B. (2015). Innovation and product innovation in marketing strategy. *Journal of Management and marketing research*, *18*, 1–25.

Kang, H. G., Mahoney, D. F., Hoenig, H., Hirth, V. A., Bonato, P., Hajjar, I. & Lipsitz, L. A. (2010). In situ monitoring of health in older adults: Technologies and issues. *Journal of the American Geriatrics Society*, *58*(8), 1579–1586.

Karrer, K., Glaser, C., Clemens, C. & Bruder, C. (2009). TA-EG Fragebogen zur Technikaffinität – Einstellung zu und Umgang mit elektronischen Geräten. *Der Mensch im Mittelpunkt technischer Systeme. 8.Berliner Werkstatt Mensch-Maschine-Systeme*, *22*(29), 196–201.

Keates, S. (2004). Developing BS7000 part 6 - Guide to managing inclusive design. In D. Stary & C. Stephandis (Eds.), *User-centered interaction paradigms for universal access in the information society. ui4all 2004. lecture notes in computer science* (Vol. 3196, pp. 332–339). Heidelberg: Springer.

Kelley, T. (2001). *The Art of Innovation: Lessons in Creativity from IDEO, America's*

Leading Design Firm. Danvers, MA: Crown Business.

Kelley, T., Woren, D. & Kelley, D. (2013). *Creative confidence* (1st ed.). New York: HarperCollins.

Kemper, S. (2012). Language in Adulthood. In F. I. M. Craik & E. Bialystok (Eds.), *Lifespan cognition: Mechanisms of change.* Oxford: Oxford University Press.

Kemper, S. & Sumner, A. (2001). The structure of verbal abilities in young and older adults. *Psychology and Aging*, *16*(2), 312–322.

Klöckner, K., Maier, A. & Nass, C. (2010). *Using Image Schemas and Their Metaphorical Extensions for Interaction Design of Intuitive- To-Use Applications – Approach and Example* (Tech. Rep.). Kaiserslautern: Fraunhofer IESE.

Knobel, M., Hassenzahl, M., Lamara, M., Sattler, T., Schumann, J., Eckoldt, K. & Butz, A. (2012). Clique Trip: feeling related in different cars. In *Proceedings of the designing interactive systems conference* (pp. 29–37).

Knoblauch, H. (1999). Metaphors, transcendence and indirect communication: Alfred Schutz'phenomenology of the life-world and the metaphors of religion. In L. Boeve, K. Frayaerts & J. Francis (Eds.), *Metaphor and god-talk* (pp. 75–94). Bern: Peter Lang.

Kobayashi, M., Hiyama, A., Miura, T., Asakawa, C., Hirose, M. & Ifukube, T. (2011). Elderly user evaluation of mobile touchscreen interactions. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque & M. Winckler (Eds.), *Human-computer interaction – interact 2011. lecture notes in computer science* (Vol. 6946, pp. 83–99). Heidelberg: Springer.

Kövecses, Z. (2006). *Metaphor in culture.* Cambridge, Massachusetts, USA: Cambridge University Press.

Kowtko, M. (2012). Using assistive technologies to improve lives of older adults and people with disabilities. In *Proceedings of systems, applications and technology conference, lisat 2012* (pp. 4–4). Long Island: IEEE.

Kuerbis, A., Mulliken, A., Muench, F., Moore, A. A. & Gardner, D. (2017). Older adults and mobile technology: Factors that enhance and inhibit utilization in the context of behavioral health. *Mental Health and Addiction Research*, *2*(2), 1–11.

Kurth, B.-M. (2013). Erste Ergebnisse aus der „Studie zur Gesundheit Erwachsener in Deutschland“ (DEGS). *Bundesgesundheitsblatt*, *55*, 980–990.

Kurzweil, R., Richter, R. & Schneider, M. L. (1990). *The age of intelligent machines.* Cambridge, Massachusetts, USA: MIT Press Cambridge.

Kyllonen, P. C. & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*(4), 389–433.

Lakoff, G., Espenson, J., Goldberg, A. & Schwartz, A. (1991). *Master Metaphor List: second draft copy.* Berkeley, CA: University of California, Cognitive Linguistics Group.

Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By* (1st ed.). Chicago, Illinois, USA: University of Chicago Press.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and* *its challenge to western thought*. New York: Basic books.

Landis, J. R. & Koch, G. G. (1977, 3). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Langdon, P. M., Lazar, J., Heylighen, A. & Dong, H. (2014). *Inclusive Designing*. Cham: Springer International Publishing.

Langdon, P. M., Lewis, T. & Clarkson, P. J. (2010). Prior experience in the use of domestic product interfaces. *Universal Access in the Information Society*, *9*(3), 209–225.

Lawry, S. G. (2012). *Identifying familiarity to facilitate intuitive interaction for older adults*. Brisbane: Queensland University of Technology.

Lee, C. & Coughlin, J. F. (2015). PERSPECTIVE: Older adults' adoption of technology: an integrated approach to identifying determinants and barriers. *Journal of Product Innovation Management*, *32*(5), 747–759.

Lee, H., Hsieh, Y.-C. & Hsu, C.-N. (2011). Adding Innovation Diffusion Theory to the Technology Acceptance Model: Supporting Employees' Intentions to use E-Learning Systems. *Educational Technology & Society*, *14*, 124–137.

LeRouge, C., Van Slyke, C., Seale, D. & Wright, K. (2014). Baby Boomers' adoption of consumer health technologies: Survey on readiness and barriers. *Journal of Medical Internet Research*, *16*(9), e200.

Lewis, T., Langdon, P. M. & Clarkson, P. J. (2008). Prior Experience of Domestic Microwave Cooker Interfaces: A User Study. In P. M. Langdon, P. J. Clarkson & P. Robinson (Eds.), *Designing inclusive futures* (pp. 95–106). London: Springer.

Lezak, M. D. (2004). *Neuropsychological assessment* (4th ed.; M. D. Lezak, D. Howleson, D. Loring, H. Hannay & J. Fischer, Eds.). New York: Oxford University Press.

Light, K. E. & Spirduso, W. W. (1990). Effects of adult aging on the movement complexity factor of response programming. *Journal of gerontology*, *45*, 107–109.

Linsey, J. S., Clauss, E. F., Kurtoglu, T., Murphy, J. T., Wood, K. L. & Markman, A. B. (2011). An experimental study of group idea generation techniques: understanding the roles of idea representation and viewing methods. *Journal of Mechanical Design*, *133*(3), 1–15.

Löffler, D. (2017). *Color, Metaphor and Culture-Empirical Foundations for User Interface Design*. Würzburg: University of Würzburg.

Löffler, D., Arlt, L., Toriizuka, T., Tscharn, R. & Hurtienne, J. (2016). Substituting Color for Haptic Attributes in Conceptual Metaphors for Tangible Interaction Design. In *Proceedings of the tei '16: Tenth international conference on tangible, embedded, and embodied interaction - tei '16* (pp. 118–125). New York: ACM.

Löffler, D., Hess, A., Hurtienne, J., Lange, K., Maier, A. & Schmitt, H. (2013). Gestaltung intuitiv benutzbarer Softwareanwendungen mit der IBIS Methode. *i-com*, *12*(2), 48–54.

Löffler, D., Hess, A., Maier, A., Hurtienne, J. & Schmitt, H. (2013). Developing intuitive user interfaces by integrating users' mental models into requirements engineering. In S. Love, K. Hone & T. McEwan (Eds.), *Proceedings of the 27th international bcs human computer interaction conference (bcs-hci '13)* (pp. 1–10). Swinton, UK: British Computer Society.

Löffler, D., Lindner, K. & Hurtienne, J. (2014). Mixing languages'. In *Proceedings of the extended abstracts of the 32nd annual acm conference on human factors in computing systems - chi ea '14* (pp. 1999–2004). New York: ACM.

Lopes, J. B. (2001). Designing user interfaces for severely handicapped persons. In *Proceedings of the 2001 ec/nsf workshop on universal accessibility of ubiquitous computing providing for the elderly - wuauc'01* (pp. 100–106). New York: ACM.

Lorenzi, N. M., Mantel, M. I. & Riley, R. T. (1912). Preparing your organization for technological change. *Healthcare informatics : the business magazine for information and communication systems*, *7*(12), 33–34.

Löwgren, J. & Stolterman, E. (2004). *Thoughtful Interaction Design.* Cambridge, MA: MIT Press Cambridge.

Lüders, M. & Brandtzæg, P. B. (2017). 'My children tell me it's so simple': A mixed-methods approach to understand older non-users' perceptions of Social Networking Sites. *new media & society*, *19*(2), 181–198.

Lund, A. (2003). *Massification of the intangible: An investigation into embodied meaning and information visualization.* Umea, Sweden: Print & Media, Umea University.

Macaranas, A., Antle, A. N. & Riecke, B. E. (2012). Bridging the gap: Attribute and Spatial Metaphors for Tangible Interface Design. In *Proceedings of the sixth international conference on tangible, embedded and embodied interaction - tei '12* (pp. 161–168). New York: ACM.

Madsen, K. H. (1994). A guide to metaphorical design. *Communications of the ACM*, *37*(12), 57–62.

Madsen, K. H. & Dalsgård, P. (2006). Inspiration card workshops. In *Proceedings of the 6th acm conference on designing interactive systems - dis '06* (pp. 2–11). New York: ACM.

Maglio, P. & Matlock, T. (1999). The conceptual structure of information space. In A. J. Munro, K. Höök & D. Benyon (Eds.), *Social navigation of information space* (1st ed., pp. 155–173). London: Springer International Publishing.

Maiden, N., Gizikis, A. & Robertson, S. (2004). Provoking creativity: Imagine what your requirements could be like. *IEEE Software*, *21*(5), 68–75.

Mandler, J. M. (2010). The spatial foundations of the conceptual system. *Language and Cognition*, *2*(1), 21–44.

Marquié, J. C., Jourdan-Boddaert, L. & Huet, N. (2002). Do older adults underestimate their actual computer knowledge? *Behaviour & Information Technology*, *21*(4), 273–280.

Martin, D. M., Laird, R., Hwang, F. & Salis, C. (2013). Computerized Short-term Memory Treatment for Older Adults with Aphasia: Feedback from Clinicians.

In *Proceedings of the 15th international acm sigaccess conference on computers and accessibility* (pp. 1–44). New York: ACM.

Matsudaira, K. & Meijer, E. (2017). Conversations with technology leaders: Erik Meijer. *Communications of the ACM*, *60*(6), 51–54.

McNab, F., Zeidman, P., Rutledge, R. B., Smittenaar, P., Brown, H. R., Adams, R. A. & Dolan, R. J. (2015). Age-related changes in working memory and the ability to ignore distraction. *Proceedings of the National Academy of Sciences*, *112*(20), 6515–6518.

Melenhorst, A.-S., Rogers, W. A. & Bouwhuis, D. G. (2006). Older adults' motivated choice for technological innovation: Evidence for benefit-driven selectivity. *Psychology and Aging*, *21*(1), 190–195.

Menard, S. (2011). Standards for standardized logistic regression coefficients. *Social Forces*, *89*(4), 1409–1428.

Merdenyan, B., Kocyigit, O., Bidar, R., Cikrikcili, O. & Salman, Y. B. (2014). Icon and User Interface Design for Mobile Banking Applications. *International Journal of Advances in Computer Science and its Applications–IJCSIA*, *4*(2), 54–59.

Mi, N., Cavuoto, L. A., Benson, K., Smith-Jackson, T. & Nussbaum, M. A. (2014). A heuristic checklist for an accessible smartphone interface design. *Universal Access in the Information Society*, *13*(4), 351–365.

Mitzner, T. L., Boron, J. B., Fausset, C. B., Adams, A. E., Charness, N., Czaja, S. J., ... Sharit, J. (2010). Older adults talk technology: Technology usage and attitudes. *Computers in Human Behavior*, *26*(6), 1710–1721.

Moneta, G. B. (2012). On the measurement and conceptualization of flow. In S. Engeser (Ed.), *Advances in flow research* (pp. 23–50). New York: Springer.

Mortensen, L., Meyer, A. S. & Humphreys, G. W. (2006). Age-related effects on speech production: A review. *Language and Cognitive Processes*, *21*(1), 238–290.

Mynatt, E. D., Essa, I. & Rogers, W. A. (2000). Increasing the opportunities for aging in place. In *Proceedings on the 2000 conference on universal usability - cuu '00* (pp. 65–71). New York: ACM.

Naumann, A. & Hurtienne, J. (2010). Benchmarks for intuitive interaction with mobile devices. In *Proceedings of the 12th international conference on human computer interaction with mobile devices and services - mobilehci '10* (pp. 401–402). New York: ACM.

Naumann, A., Hurtienne, J., Israel, J. H., Mohs, C., Kindsmüller, M. C., Meyer, H. A. & Hußlein, S. (2007). Intuitive use of user interfaces: Defining a vague concept. *Engineering Psychology and Cognitive Ergonomics*, 128–136.

Neale, D. C. & Carroll, J. M. (1997). The role of metaphors in user interface design. In *Handbook of human-computer interaction* (2nd ed., pp. 441–462). Amsterdam, Netherlands: Elsevier.

Neves, B. B., Franz, R. L., Munteanu, C., Baecker, R. & Ngo, M. (2015). "My Hand Doesn't Listen to Me!": Adoption and Evaluation of a Communication Technology for the 'Oldest Old'. In *Proceedings of the acm chi'15 conference on human factors in computing systems* (Vol. 1, pp. 1593–1602). New York: ACM.

Nicolau, H., Guerreiro, T., Jorge, J. & Gonzalves, D. (2014). Mobile touchscreen user interfaces: Bridging the gap between motor-impaired and able-bodied users. *Universal Access in the Information Society*, *13*(3), 303–313.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 152–158). New York: ACM.

Norman, D. A. (2013). *The design of everyday things: Revised and expanded edition.* New York: Basic Books (AZ).

Norman, D. A. & Verganti, R. (2014). Incremental and Radical Innovation: Design Research vs. Technology and Meaning Change. *Design Issues*, *30*(1), 78–96.

Nurgalieva, L., Laconich, J. J. J., Baez, M., Casati, F. & Marchese, M. (2017). Designing for older adults: review of touchscreen design guidelines. *arXiv preprint arXiv:1703.06317*.

O'Brien, M. A. (2010). *Understanding human-technology interactions: The role of prior experience and age (Doctoral Dissertation)* (No. May). Atlanta, Georgia: Georgia Institute of Technology.

Olson, K. E., O'Brien, M. A., Rogers, W. A. & Charness, N. (2011). Diffusion of technology: Frequency of use for younger and older adults. *Ageing International*, *36*(1), 123–145.

Owen, C. L. (2006). *Design Thinking : Driving Innovation.* Retrieved from `BPMInstitute.org`

Page, T. (2014). Touchscreen mobile devices and older adults: a usability study. *International Journal of Human Factors and Ergonomics*, *3*(1), 65–85.

Park, D. C. & Festini, S. B. (2016). Theories of Memory and Aging: A Look at the Past and a Glimpse of the Future. *The journals of gerontology. Series B, Psychological sciences and social sciences*, *72*(1), 82–90.

Peace, S. (2016). Age-Inclusive Design: A Challenge for Kitchen Living? In R. Imrie (Ed.), *Care and design* (pp. 18–35). Hoboken, New Jersey, USA: John Wiley & Sons.

Pew Research Center. (2017). *Tech Adoption Climbs Among Older Adults.* Retrieved from `http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/`

Pichora-Fuller, M. K., Schneider, B. A., Benson, N. J., Hamstra, S. J. & Storzer, E. (2006). Effect of age on detection of gaps in speech and nonspeech markers varying in duration and spectral symmetry. *The Journal of the Acoustical Society of America*, *119*(2), 1143.

Plett, H. F. (1983). *Einführung in die rhetorische Textanalyse.* Hamburg: Buske Verlag.

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W. & Ferri, C. P. (2013). The global prevalence of dementia: A systematic review and metaanalysis. *Alzheimer's and Dementia*, *9*(1), 63–75.

Rabin, L. A., Barr, W. B. & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65.

Rabitt, P. (1965). An age-decrement in the ability to ignore irrelevant information. *Journal of Gerontology*, *20*, 233–238.

Radford, S. K. & Bloch, P. H. (2011). Linking innovation to design: Consumer responses to visual product newness. *Journal of Product Innovation Management*, *28*(S1), 208–220.

Radvansky, G. A. & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic bulletin & review*, *14*(6), 1027–1042.

Raju, P. S. (1980). Optimum Stimulation Level: Its Relationship to Personality, Demographics, and Exploratory Behavior. *Journal of Consumer Research*, *7*(3), 272–282.

Rasmusson, X. D., Zonderman, A. B., Kawas, C. & Resnick, S. M. (1998). Effects of Age and Dementia on the Trail Making Test. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D)*, *12*(2), 169–178.

Raubal, M., Egenhofer, M., Pfoser, D. & Tryfona, N. (1997). Structuring Wayfinding Tasks with Image Schemata: Wayfinding in Airport as a Case Study. In S. Hirtle & A. Frank (Eds.), *Proceedings of cosit'97* (pp. 85–102). Laurel Highlands, PA: Springer Verlag.

Raubal, M. & Worboys, M. (1999). A formal model of the process of wayfinding in built environments. *Spatial information theory. Cognitive and Computational Foundations of Geographic Information Science*, *1661*, 381–399.

Razak, F. & Razak, N. (2013). How Simple is Simple: Our Experience with Older Adult Users. In *Proceedings of the india hci 2013 and apchi, 2013* (pp. 379–387). New York: ACM.

Reason, J. (1990). *Human error*. Cambridge, Massachusetts, USA: Cambridge university press.

Reddy, G. R. (2012). *Approaches to designing for older adults' intuitive interaction with complex devices (Doctoral Dissertation)*. Queensland: Queensland University of Technology.

Reddy, G. R., Blackler, A., Popovic, V. & Mahar, D. P. (2014). Adaptable interface as an approach to address diversity in older users' capabilities. In Y.-K. Lim, K. Neidderer, J. Redström, E. Stolterman & A. Valtonen (Eds.), *Drs 2014 : Design's big debates pushing the boundaries of design research* (pp. 374–387). Umea, Sweden: Umea Institute of Design, Umea University.

Reeves, W. W. (1999). *Learner-centered Design : A Cognitive View of Managing Complexity in Product, Information, and Environmental Design*. Thousan Oaks, CA: Sage Publications.

Reio, T. G. & Choi, N. (2004). Novelty Seeking in Adulthood: Increases Accompany Decline. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *165*(2), 119–133.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, *8*, 271–276.

Reitan, R. M. & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review*, *4*(3), 161–198.

Rickards, T. (1999). Brainstorming. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (pp. 219–228). San Diego: Academic Press San Diego, CA.

Rittel, H. W. J. & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169.

Roehrich, G. (1995). Innvoativités hédoniste et sociale: proposition d'une échelle de mesure. *Recherche et Applications en Marketing*, *9*(2), 19–42.

Roehrich, G. (2004). Consumer innovativeness - Concepts and measurements. *Journal of Business Research*, *57*(6), 671–677.

Rogers, E. M. (1995). *Diffusion of innovations* (3rd ed.). New York: Macmillan Publishing Co.

Rogers, W. A., Meyer, B., Walker, N. & Fisk, A. D. (1998). Functional limitations to daily living tasks in the aged: a focus group analysis. *Human factors*, *40*(1), 111–125.

Rogers, Y., Sharp, H. & Preece, J. (2011). *Interaction design: beyond human-computer interaction*. Hoboken, New Jersey, USA: John Wiley & Sons.

Romano Bergstrom, J. C., Olmsted-Hawala, E. L. & Bergstrom, H. C. (2016). Older adults fail to see the periphery in a Web site task. *Universal Access in the Information Society*, *15*(2), 261–270.

Romano Bergstrom, J. C., Olmsted-Hawala, E. L. & Jans, M. E. (2013). Age-related differences in eye tracking and usability performance: website usability for older adults. *International Journal of Human-Computer Interaction*, *29*(8), 541–548.

Roque, N. A. & Boot, W. R. (2016). A new tool for assessing mobile device proficiency in older adults: the mobile device proficiency questionnaire. *Journal of Applied Gerontology*, *37*(2), 131–156.

Saffer, D. (2005). *The role of metaphor in interaction design (Doctoral Dissertation)*. Pittsburgh.

Saffer, D. (2010). *Designing for interaction: creating innovative applications and devices*. San Francisco, CA: New Riders.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 400–426). New York: Van Nostrand Reinhold Company.

Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology*, *8*(4), 535–543.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403–428.

Salthouse, T. A. (2010). Relations between Age and Cognitive Functioning. In T. A. Salthouse (Ed.), *Major issues in cognitive aging* (pp. 3–34). Oxford:

Oxford University Press.

Salthouse, T. A. & Babcock, R. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–777.

Salthouse, T. A. & Craik, F. I. M. (2000). Closing Comments. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 689–703). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Sanchiz, M., Chevalier, A., Fu, W.-T. & Amadieu, F. (2017). Relationships between Age, Domain Knowledge and Prior Knowledge Pre-activation on Information Searching. In *Proceedings of the 2017 conference on conference human information interaction and retrieval* (pp. 289–292). ACM.

Sauro, J. & Dumas, J. (2009). Comparison of Three One-Question, Post-Task Usability Questionnaires. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1599–1608.

Schmidt, K. H. & Metzler, P. (1992). *WST–Wortschatztest.* Weinheim: Beltz.

Schmidt, L. C., Vargas-Hernandez, N., Kremer, G. & Linsey, J. S. (2010). Pilot of systematic ideation study with lessons learned. In *international design engineering technical conferences & computers and information in engineering conference asme 2010* (pp. 771–779). New York: ASME.

Schmitt, R. (1983). Systematic Metaphor Analysis as a Method of Qualitative Research. *The Qualitative Report*, *10*, 358–394.

Schmitt, R. (2003). Methode und Subjektivität in der Systematischen Metaphernanalyse. *Forum Qualitative Sozialforschung*, *4*(2), Art. 41.

Schneider, B. A., Daneman, M. & Murphy, D. R. (2005). Speech Comprehension Difficulties in Older Adults: Cognitive Slowing or Age-Related Changes in Hearing? *Psychology and Aging*, *20*(2), 261–271.

Schumpeter, J. A. (1939). *Business cycles: a theoretical, historical, and statistical analysis of the capitalist process.* New York: McGraw-Hill.

Sengpiel, M. (2016). Teach or design? how older adults' use of ticket vending machines could be more effective. *ACM Transactions on Accessible Computing (TACCESS)*, *9*(1), 1–27.

Sengpiel, M. & Dittberner, D. (2008). The computer literacy scale (CLS) for older adults - development and validation. In M. Herczeg & M. C. Kindsmüller (Eds.), *Mensch & computer 2008: Viel mehr interaktion* (pp. 7–17). München: Oldenbourg Wissenschaftsverlag.

Sethi, R., Smith, D. C. & Park, C. W. (2001). Cross-functional product development teams, creativity, and the innovativeness of new consumer products. *Journal of marketing research*, *38*(1), 73–85.

Shake, M. C., Noh, S. R. & Stine-Morrow, E. A. L. (2009). Age differences in learning from text: Evidence for functionally distinct text processing systems. *Applied Cognitive Psychology*, *23*(4), 561–578.

Shneiderman, B. (2003). Leonardo's Laptop: Human Needs and the New Computing Technologies. *The Electronic Library*, *21*, 174–175.

Silva, P. A. & Read, J. C. (2010). A methodology to evaluate creative design methods: a study with the BadIdeas method. In *Proceedings of the 22nd conference of the computer-human interaction special interest group of australia on computer-human interaction* (pp. 264–271). ACM.

Slappendel, C. (1996). Perspectives on Innovation in Organizations. *Organization Studies*, *17*(1), 107–129.

Slepian, M. L. & Ambady, N. (2014). Simulating sensorimotor metaphors: Novel metaphors influence sensory judgments. *Cognition*, *130*(3), 309–314.

Smiler, A. P., Gagne, D. D. & Stine-Morrow, E. a. L. (2003). Aging, memory load, and resource allocation during reading. *Psychology and aging*, *18*(2), 203–209.

Smith, A. (2014). *Older Adults and Technology Use* (Tech. Rep. No. April). Washington, DC, US: Pew Research Center.

Smith, S. L. & Mosier, J. N. (1986). *Guidelines for Designing User Interface Software.* Bedford, MA: The MITRE Corporation.

Solon, O. (2017). *Killer robots? Musk and Zuckerberg escalate row over dangers of AI.* Retrieved from `https://www.theguardian.com/technology/2017/jul/25/elon-musk-mark-zuckerberg-artificial-intelligence-facebook-tesla`

Sonderegger, A., Schmutz, S. & Sauer, J. (2016). The influence of age in usability testing. *Applied Ergonomics*, *52*, 291–300.

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A., Krennmayr, T. & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU* (Vol. 14). Amsterdam, Netherlands: John Benjamins Publishing.

Steenkamp, J.-B. E. M., Hofstede, F. & Wedel, M. (1999). A Cross-National Investigation into the Individual and National Cultural Antecedents of Consumer Innovativeness. *Journal of Marketing*, *63*(2), 55–69.

Stefanowitsch, A. (2006). Words and their metaphors: A corpus-based approach. *Corpus-based Approaches to Metaphor and Metonymy*(1), 61–105.

Stokes, D. E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation* (Vol. 92) (No. 3). Washington, DC, US: Brookings Institution Press.

Story, M. F., Mueller, J. L. & Mace, R. L. (1998). *The universal design file: Designing for people of all ages and abilities.* Raleigh, North Carolina, USA: NC State University.

Struve, D. (2010). *Instruktionsdesign für ältere Nutzer interaktiver Systeme: Gestaltungsaspekte modellbasierter Lernvideos in multimedialen Bedientrainings.* Logos-Verlag.

Struve, D. & Wandke, H. (2009). Video modeling for training older adults to use new technologies. *ACM Transactions on Accessible Computing (TACCESS)*, *2*(1), 4.

Stuart-Hamilton, I. (2002). *The psychology of ageing: An introduction.* London: Jessica Kingsley Publishers.

Tabachnick, B. G., Fidell, L. S. & others. (2001). *Using multivariate statistics.* London: Pearson Education.

Thornton, R. & Light, L. L. (2006). Language Comprehension and Production in Normal Aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology* *of aging* (pp. 261–287). Burlington, Massachusetts, USA: Elsevier.

Trotter, R. T. (2012). Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives. *Preventive Medicine*, *55*(5), 398–400.

Tscharn, R. (2017). Design of Age-Inclusive tangible User interfaces using image-Schematic metaphors. In *Proceedings of the 11th international conference on tangible, embedded, and embodied interaction (tei)* (pp. 693–696). New York: ACM.

Tscharn, R., Ly-Tung, N., Löffler, D. & Hurtienne, J. (2016). Ambient light as spatial attention guidance in indoor environments. *UbiComp 2016 Adjunct - Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 1627–1630.

Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. New York: Basic Books.

United Nations, D. o. E. & Affairs, S. (2017). *World Population Prospects: The 2017 Revision. Key Findings and Advance Tables* (Vol. 2). New York.

van der Elst, W., van Boxtel, M. P. J., van Breukelen, G. J. P. & Jolles, J. (2006). The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. *Journal of clinical and experimental neuropsychology*, *28*(6), 998–1009.

van de Sand, F. (2017). *User Experience Identity Mit Neuropsychologie digitale Produkte zu Markenbotschaftern machen* (1st ed.). Berlin: Springer Gabler.

Van Raaij, E. M. & Schepers, J. J. L. (2008). The acceptance and use of a virtual learning environment in China. *Computers & Education*, *50*(3), 838–852.

Venkatesan, M. (1973). Cognitive consistency and novelty seeking. In S. Ward & T. S. Robertson (Eds.), *Consumer behavior–theoretical sources* (pp. 355–384). Englewood Cliffs: Prentice Hall.

Venkatesh, V. & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management science*, *46*(2), 186–204.

Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, *27*(3), 425–478.

Vercruyssen, M. (1997). Movement control and speed of behavior. In W. A. Rogers & A. D. Fisk (Eds.), *Handbook of human factors and the older adult* (pp. 55–86). San Diego, CA: Academic Press.

Vesneski, M., Schmitter-Edgecombe, M. & Jones, D. (1998). Aging and Word Finding: A Comparison of Discourse and Non-Discourse Tests. *Archives of Clinical Neuropsychology*, *13*(1), 46.

Vines, J., Blythe, M., Dunphy, P. & Monk, A. (2011). Eighty something: banking for the older old. In *Proceedings of the 25th bcs conference on human-computer interaction* (pp. 64–73). Swindon: British Computer Society.

Vines, J., Blythe, M., Lindsay, S., Dunphy, P., Monk, A. & Olivier, P. (2012). Questionable Concepts: Critique as a Resource for Designing with Eighty Somethings. In *Proceedings of the 2012 conference on human factors in computing systems - chi'12* (pp. 1169–1178). New York: ACM.

Vines, J., Pritchard, G., Wright, P. & Olivier, P. (2015). An Age-Old Problem : Examining the Discourses of Ageing in HCI and Strategies for Future Research. *Transactions on Computer-Human Interaction*, *22*(1), 1–27.

Vogt, D. (2013). *Innovation Perception from a Customer Perspective (Doctoral Dissertation).* St. Gallen: University of St. Gallen.

Vorrink, S. N. W., Antonietti, A. M., Kort, H. S. M., Troosters, T., Zanen, P. & Lammers, J.-W. J. (2017). Technology use by older adults in the Netherlands and its associations with demographics and health outcomes. *Assistive Technology*, *29*(4), 188–196.

Vroman, K. G., Arthanat, S. & Lysack, C. (2015). "Who over 65 is online?" Older adults' dispositions toward information communication technology. *Computers in Human Behavior*, *43*, 156–166.

Wagner, N., Hassanein, K. & Head, M. (2010). Computer use by older adults: A multi-disciplinary review. *Computers in human behavior*, *26*(5), 870–882.

Wagner, N., Hassanein, K. & Head, M. (2014). The impact of age on website usability. *Computers in Human Behavior*, *37*, 270–282.

Walker, N., Philbin, D. A. & Fisk, A. D. (1997). Age-Related Differences in Movement Control: Adjusting Submovement Structure To Optimize Performance. *Journal of Gerontology: Psychological Sciences*, *52*(I), 4–52.

Walsh, K. & Callan, A. (2011). Perceptions, Preferences, and Acceptance of information and communication technologies in older-adult community care settings in Ireland: A case-study and ranked-care program analysis. *Ageing International*, *36*(1), 102–122.

Wandke, H., Sengpiel, M. & Sönksen, M. (2012). Myths about older people's use of information and communication technology. *Gerontology*, *58*(6), 564–570.

Watson, D., Clark, L. A. & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063.

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale.* Oxford: Psychological Corporation.

Wechsler, D. (1997). *WAIS-III administration and scoring manual.* San Antonio, TX: Psychological Corporation, Harcourt Brace & Co.

Weizenbaum, J. (1972). On the Impact of the Computer on Society. *Science*, *176*(4035), 609–614.

Wilkie, K., Holland, S. & Mulholland, P. (2010). What can the language of musicians tell us about music interaction design? *Computer Music Journal*, *34*(4), 34–48.

Wingfield, A., Lindfield, K. C. & Harold, G. (2000, 8). Effects of age and hearing sensitivity on the use of prosodic information in spoken word recognition.

Journal of Speech, Language, and Hearing Research, *43*(4), 915–925.

Wingfield, A., Stine, E. a., Lahar, C. J. & Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Experimental aging research*, *14*(2-3), 103–107.

Winkler, A., Baumann, K., Huber, S., Tscharn, R. & Hurtienne, J. (2016). Evaluation of an Application Based on Conceptual Metaphors for Social Interaction Between Vehicles. In *Proceedings of the 2016 acm conference on designing interactive systems - dis '16* (pp. 1148–1159). New York: ACM.

Worinkeng, E., Summers, J. D. & Joshi, S. (2013). Can a pre-sketching activity improve idea generation? *Smart Product Engineering*, 583–592.

Zaphiris, P., Sustar, H. & Pfeil, U. (2008). *Inclusive Design for Older People* (Tech. Rep.). London: Centre for HCI Design, City University London.

Zhang, S., Grenhart, W. C. M., McLaughlin, A. C. & Allaire, J. C. (2017). Predicting computer proficiency in older adults. *Computers in Human Behavior*, *67*, 106–112.

Ziefle, M. & Bay, S. (2005). How older adults meet complexity: Aging effects on the usability of different mobile phones. *Behaviour & Information Technology*, *24*(5), 375–389.

Zijlstra, F. R. H. (1993). *Efficiency in work behaviour: A design approach for modern tools (Doctoral Dissertation)*. Delft: Delft University Press.

Zimmer, Z. & McDaniel, S. A. (2016). *Global Ageing in the Twenty-First Century: Challenges, Opportunities and Implications*. New York: Routledge.

Zimmerman, J. & Forlizzi, J. (2014). Research through design in HCI. In J. Olson & W. Kellogg (Eds.), *Ways of knowing in hci* (1st ed., pp. 167–189). New York: Springer.

Zimmerman, J., Forlizzi, J. & Evenson, S. (2007). Research Through Design as a Method for Interaction Design Research in HCI design research in HCI. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 493–502). New York: ACM.

Zuckerman, M. (1979). Sensation seeking: Beyond the optimal level of arousal. *Personality and Individual Differences*, *1*(2), 191–193.

Zuckerman, M., Eysenck, S. & Eysenck, H. J. (1978). Sensation Seeking in England and America: Cross-cultural, Age, and Sex Comparisons. *Journal of Consulting and Clinical Psychology*, *46*(1), 139–149.

Appendix A

Appendix

A.1 Appendix Study 1

A.1.1 Instruction of Abstract Concepts

In eigenen Worten, max. 1-2 Sätze... (was verstehen Sie genau damit?)

Domäne Geldverwaltung

1. Was ist „Geld“?
2. Was ist eine Bank?
3. Was ist eine Überweisung? Was passiert dabei?
4. Wenn Sie eine Überweisung tätigen wollen, was müssen Sie dafür tun?
5. Was sind Zinsen?
6. Warum bekommt man die?
7. Was ist Sparen?
8. Was ist ein Dauerauftrag? Was passiert dabei?
9. Was müssen Sie für einen neuen Dauerauftrag tun?
10. Was ist ein Konto? Was kann man damit machen?
11. Was ist, wenn der Kontostand negativ ist?
12. Warum haben manche Menschen mehrere Konten?
13. Was sind Kundenberater?

Andere

1. Was ist eine Erinnerung?
2. Was ist eine Bestellung?
3. Was ist eine Reservierung?
4. Was ist eine Empfehlung?
5. Was ist ein Telefonat?
6. Was ist eine Begrüßung?
7. Was ist Freundschaft?
8. Was ist ein Abenteuer?
9. Was ist eine Unterhaltung?
10. Was ist Energie?
11. Was ist eine Erkenntnis?
12. Was tut man in einem Gespräch?
13. Was ist Lernen?

A.1.2 Demographic questionnaire and prior technological knowledge

VP-Code: ____________________________ Datum und Uhrzeit: _________________________

Bitte beantworten Sie alle Fragen. Die spontane Antwort ist dabei meist die richtige. Sollten Sie eine Frage nicht beantworten oder eine bereits gesetzte Antwort ändern wollen, so teilen Sie dies bitte der Versuchsleiterin oder dem Versuchsleiter mit (VP-Code: Jeweils Erste und zweite Stelle des Vornamens der Mutter und des Vater + Geburts*tag.* Z.B.: MAPE22)

Alter: __________________

Geschlecht: __________________

Höchster Schulabschluss: __________________

Welche der folgenden Geräte besitzen Sie schon wie lange (in Jahren)? Nicht zutreffende bitte streichen.

Smartphone ____	Mobiltelefon (mit Tasten) ____	Navigationsgerät ____
Tablet ____	Laptop / Notebook ____	Desktop-PC ____

Wie oft haben Sie folgende Tätigkeiten in den letzten 12 Monaten an einem Gerät mit Bildschirm (z.B. Tablet, Smartphone, PC) durchgeführt?

	nie	seltener als einmal pro Monat	1-2 mal pro Monat	1-2 mal pro Woche	3-4 mal pro Woche	täglich
E-Mail geschrieben	○	○	○	○	○	○
E-Mail gelesen	○	○	○	○	○	○
SMS verschickt	○	○	○	○	○	○
SMS gelesen	○	○	○	○	○	○
Instant-Messenger genutzt (z.B. Whatsapp, Telegram)	○	○	○	○	○	○
Kalender genutzt	○	○	○	○	○	○
Telefoniert	○	○	○	○	○	○
Zeitung gelesen	○	○	○	○	○	○
Online-Shopping	○	○	○	○	○	○
Social Media	○	○	○	○	○	○
Text bearbeitet	○	○	○	○	○	○
Bilder bearbeitet	○	○	○	○	○	○
Spiele gespielt	○	○	○	○	○	○
Fotografiert	○	○	○	○	○	○
Musik gehört	○	○	○	○	○	○
Fahrkarte gekauft	○	○	○	○	○	○
Bus- / Zugverbindung abgerufen	○	○	○	○	○	○
Geld abgehoben (Automaten)	○	○	○	○	○	○
Geld überwiesen (am Automaten)	○	○	○	○	○	○
Kontostand abgefragt (am Automaten)	○	○	○	○	○	○

Bitte wenden!

VP-Code: _____________

Wie oft haben Sie diese Geräte in Ihrem Leben bereits genutzt?

	nie	einmal	selten	mehrmals	regelmäßig
Smartphone	○	○	○	○	○
Tablet	○	○	○	○	○
Notebook / Laptop	○	○	○	○	○
Desktop-PC	○	○	○	○	○
Fahrkartenautomat	○	○	○	○	○
Spielautomat	○	○	○	○	○
Fernseher	○	○	○	○	○
Normales Telefon	○	○	○	○	○
Bankautomat	○	○	○	○	○
Tastatur	○	○	○	○	○
Maus	○	○	○	○	○
Tracking-Pad	○	○	○	○	○
MP3-Player / Disc-Man	○	○	○	○	○
Getränkeautomat	○	○	○	○	○
Infotainment-System	○	○	○	○	○

➔ STOP: Interviewteil

Wie oft benutzen Sie folgende Arten des Banking?

	nie	seltener als einmal pro Monat	1-2 mal pro Monat	1-2 mal pro Woche	3-4 mal pro Woche	täglich
Am Automaten	○	○	○	○	○	○
Am Schalter	○	○	○	○	○	○
Am Computer	○	○	○	○	○	○
Am Tablet	○	○	○	○	○	○
Am Smartphone	○	○	○	○	○	○
Am Telefon	○	○	○	○	○	○
Am Smartphone (Internetseite)	○	○	○	○	○	○
Am Smartphone (App)	○	○	○	○	○	○

Welche Funktionen haben Sie schon einmal beim Online-Banking benutzt?

Online-Überweisung mit TAN-Liste	○	Online-Überweisung mit chipTAN-Gerät	○	Online-Überweisung mit SMS-TAN	○
Terminüberweisung	○	Kundenberater angeschrieben	○	Finanzstatus überprüft	○
Handy aufgeladen	○	iTunes aufgeladen	○	Kreditkarte überprüft	○
Über Angebote informiert	○	Dauerauftrag eingerichtet	○	Umbuchung	○
Auslandsüberweisung	○	Lastschrift zurückgegeben	○	Karte gesperrt	○
Freistellungsauftrag	○	Persönliche Daten geändert (z.B. Adresse)	○	Eingeloggt	○
Kontoauszüge abgerufen	○	Kreditkartenrahmen verändert	○	Disporahmen geändert	○

Fragebogen – v2.0 2

VP-Code: _____________

Wie gut schätzen Sie sich selbst beim Umgang mit Online-Banking ein?

Sehr schlecht ---	--	-	Neutral o	+	++	Sehr gut +++
○	○	○	○	○	○	○

Wie gut treffen folgende Aussagen auf Sie zu?

	Trifft gar nicht zu 1	Trifft eher nicht zu 2	Teils/ teils 3	Trifft eher zu 4	Trifft voll zu 5
1. Ich liebe es, neue elektronische Geräte zu besitzen.	○	○	○	○	○
2. Elektronische Geräte machen krank.	○	○	○	○	○
3. Ich gehe gern in den Fachhandel für elektronische Geräte.	○	○	○	○	○
4. Ich habe bzw. hätte Verständnisprobleme beim Lesen von Elektronik-und Computerzeitschriften.	○	○	○	○	○
5. Elektronische Geräte ermöglichen einen hohen Lebensstandard.	○	○	○	○	○
6. Elektronische Geräte führen zu geistiger Verarmung.	○	○	○	○	○
7. Elektronische Geräte machen vieles umständlicher.	○	○	○	○	○
8. Ich informiere mich über elektronische Geräte, auch wenn ich keine Kaufabsicht habe.	○	○	○	○	○
9. Elektronische Geräte machen unabhängig.	○	○	○	○	○
10. Es macht mir Spaß, ein elektronisches Gerät auszuprobieren.	○	○	○	○	○
11. Elektronische Geräte erleichtern mir den Alltag.	○	○	○	○	○
12. Elektronische Geräte erhöhen die Sicherheit.	○	○	○	○	○
13. Elektronische Geräte verringern den persönlichen Kontakt zwischen den Menschen.	○	○	○	○	○
14. Ich kenne die meisten Funktionen der elektronischen Geräte, die ich besitze.	○	○	○	○	○
15. Ich bin begeistert, wenn ein neues elektronisches Gerät auf den Markt kommt.	○	○	○	○	○
16. Elektronische Geräte verursachen Stress.	○	○	○	○	○
17. Ich kenne mich im Bereich elektronischer Geräte aus.	○	○	○	○	○
18. Es fällt mir leicht, die Bedienung eines elektronischen Geräts zu lernen.	○	○	○	○	○
19. Elektronische Geräte helfen, an Informationen zu gelangen.	○	○	○	○	○

➔ STOP: CLS

Fragebogen – v2.0 3

A.2 Appendix Study 2

A.2.1 Contextual Interviews: Basic questions.

Gefahrenes Auto / Fahrleistung pro Jahr / Tätigkeit / Alter / Smartphone Nutzung:

Im Auto vor mir… (Vernetztes Infotainment)

- Als Sie das letzte Mal mit dem Auto in den Urlaub gefahren sind, was haben Sie neben der Fahraufgabe noch getan? (welche Medien genutzt, Gespräche mit Mitfahrern, Pausen an speziellen Orten)
- Als Sie das letzte Mal eine längere Überland-Fahrt gefahren sind, was haben Sie neben der Fahraufgabe noch getan?
- Welche elektronischen Geräte verwenden Sie in ihrem Auto? Wie nutzen Sie diese?
- Unterschiede zwischen Stadt und Autobahnfahrt?
- Unterschiede zwischen beruflichen und privaten Fahrten?
- Wann wollten Sie zuletzt während einer Fahrt einer anderen Person –die nicht im Auto saß - etwas mitteilen. (Kommunikation mit Zentrale, anderen Fahrern?) Z.B. Wann fuhren Sie zuletzt mit Freunden in verschiedenen Autos zum gleichen Ziel (als Kolonne)?
- Wie kommunizierten Sie mit ihnen während der Fahrt? Kolonnenfahrten auch mit mehreren Taxis zum gleichen Ziel: kommunizieren Fahrer oder Fahrgäste Fahrzeugübergreifend?
- Welche schönen sozialen Erfahrungen haben Sie im Auto bisher gemacht?
- Mehrwert durch Kommunikation mit anderen Autos/Fahrern?
- Welche neuen Funktionen können Sie sich vorstellen?
- Wo im Auto sollte das sein?
- Wie interagiert man damit? zur Interaktion auffordern!
- Auch/nur für den Beifahrer?
- Als Sie zuletzt ein Lied gehört/einen Film gesehen haben, der so gut war, dass Sie ihn einem Freund empfehlen wollten, der gerade nicht dabei war, wie haben Sie das gemacht?
- Potentielle Funktionen aufzählen, diskutieren.
- Mit Interviewpartner Ideen generieren und etwas rumspinnen
- 5 use cases, rumspielen lassen
- Eventuell kleine Fokusgruppen, gerade mit älteren Nutzern? (LKW-Fahrer an Raststätte…)
- Für Prototyping: Autobahn, Rastplatz, innerorts? Kann man überholen? Wie sieht die Landschaft aus? Abstraktes Bild vom Fahrzeug mit Personen, Blick auf Straße, Minimale Kantendarstellung, Landschaft.

A.3 Appendix Study 3

A.3.1 Use-Case Instruction

1. **Use-Case 1**: Es ist wieder soweit! Sie haben eine neue Wohnung un müssen zuerst die Kaution auf das Konto des Vermieters überweisen. Dafür loggen Sie sich in Ihre App ein, geben alle benötigten Daten ein und überweisen den Betrag von 900€ auf das Kautionskonto des Vermieters.

   Um die Miete rechtzeitig jeden Monat zu überweisen, richten Sie auch einen Dauerauftrag für den betrag von 340€ ein. Dieser wird immer am Monatsanfang auf ein weiteres Konto des Vermieters überwiesen.

   Um Ihre aktuellen Finanzstatus zu sehen, kehren Sie auf die Startseite zurück und loggen sich nach einer kurzen Kontrolle aus.

2. **Use-Case 2**: Sie besitzen mehrere Konten bei unterschiedlichen Banken. Da jede Bank eine eigene Seite für das Onlinebanking bereitstellt, ist die Geldverwaltung sehr umständlich. Jedoch besitzen Sie eine App, die all diese Konten zusammenführt.

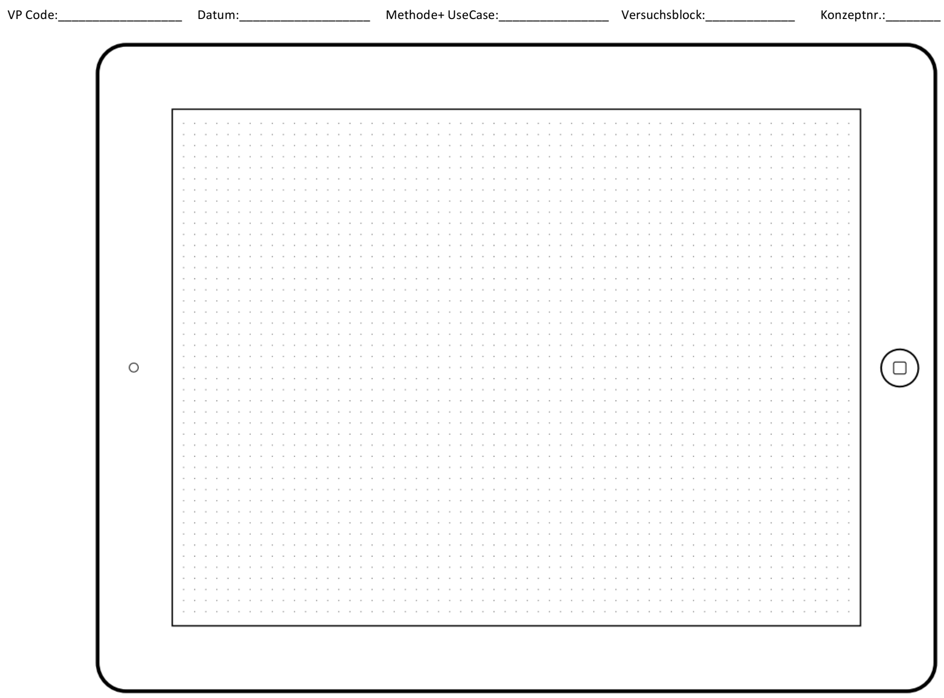
   Heute fügen Sie ein neues Konto zu dieser App hinzu und zwar das Konto Ihrer Tochter. Um die Überweisung des Geburtstaggeldes nicht zu vergessen, verschieben Sie den Betrag von 300€ direkt auf das neu angelegte Konto.

   Um festzustellen, wie viele Zinsen das neu angelegte Konto abwirft, sehen Sie sich die Zinsprognose für das kommende Jahr an.

Note: participants created a paper-prototype (tablet-size) that should be able to (depending on the experimental condition) enable a user to complete one use-case.

A.3.2 Design Template

Note: the size in the experiment was DinA4.



A.3.3 Questionnaire used before the study

VP-Code:________________________ (**Vo**rnameMutter, **Vo**rnameVater, **Geburtsmonat**: z.B. MaPe05)

Datum:__________________ Team:_______ Bedingung:__________________

Demographische Fragen und Vorerfahrung

Alter: _______

Geschlecht: □ weiblich □ männlich

Studiengang: ___________________________________

Semester: _____ (z.B.: 1, für 1.Semester)

Wie sehr interessiert Sie das Thema Interaktionsdesign?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr stark

In wie vielen Projekten haben Sie bisher Techniken der benutzerzentrierten Gestaltung praktisch angewendet (z.B. Affinity Diagramme, Papierprototypen, Design Thinking, Image Schemas, etc.)?

□ (0) □ (1) □ (2) □ (3) □ (4) □ (5) □ (mehr als 5 Projekte)

Welche Methoden waren das?

Ihrer persönlichen Einschätzung nach: wie viel Erfahrung haben Sie mit Interaktionsdesign?

keine □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr viel

Haben Sie Erfahrung mit **Tools** wie Axure, Photoshop CS, InDesign, etc? Wenn ja, mit welchen?

Wie oft benutzen Sie Online-Banking?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr viel

Wenn ja, für welche Aufgaben und wieso?

A.3.4 Questionnaire used after the study

VP-Code:_________________________ (**Vo**rnameMutter, **Vo**rnameVater, **Geburtsmonat**: z.B. MaPe05)

Datum:___________________ Team:_______ Bedingung:___________________

Nachbefragung - Gesamt

In diesem Fragebogen sollen Sie einige abschließende Fragen zum Versuch beantworten.

Was war das Ziel der Studie?

Wie gut hat Ihrer Meinung nach das Teamwork funktioniert?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Anmerkungen:

Wie gut kannten Sie die Methode „Affinity Diagramme“ bereits?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Wie bewerten Sie die Qualität des Affinity Diagramms?

sehr schlecht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Anmerkungen:

Wie gut sind Sie mit dem Design-Prozess zurechtgekommen?

sehr schlecht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Wie leicht war die Methode anzuwenden?

sehr schlecht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Wie strukturiert war der Designprozess insgesamt?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Wie sehr hat diese Methode Ihre Kreativität angeregt?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Hatten Sie eher wenige, aber durchdachte, oder eher viele spontane Ideen?

durchdachte □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) spontane

Wie hilfreich war die Methode für den Designprozess?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Wie gut war der Use Case zu bearbeiten hinsichtlich der Entwicklung von Ideen?

sehr schlecht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

1

VP-Code:________________________ (**Vo**rnameMutter, **Vo**rnameVater, **Geburtsmonat**: z.B. MaPe05)

Datum:__________________ Team:_______ Bedingung:__________________

Was war heute besonders einfach am Versuch (generell)?

Was war heute besonders schwer am Versuch (generell)?

Geben Sie an, wie Sie sich in Bezug auf das Erstellen der Prototypen gefühlt haben

	Niedrig									**Hoch**
Herausforderungen der Tätigkeit	0	1	2	3	4	5	6	7	8	9
Ihre Fähigkeiten für die Tätigkeit	0	1	2	3	4	5	6	7	8	9
War diese Tätigkeit wichtig für Sie?	0	1	2	3	4	5	6	7	8	9
Waren Sie zufrieden mit dem, was Sie erreicht haben?	0	1	2	3	4	5	6	7	8	9

Wie schätzen Sie selbst Ihre Prototypen-Skizzen hinsichtlich folgender Kriterien ein?

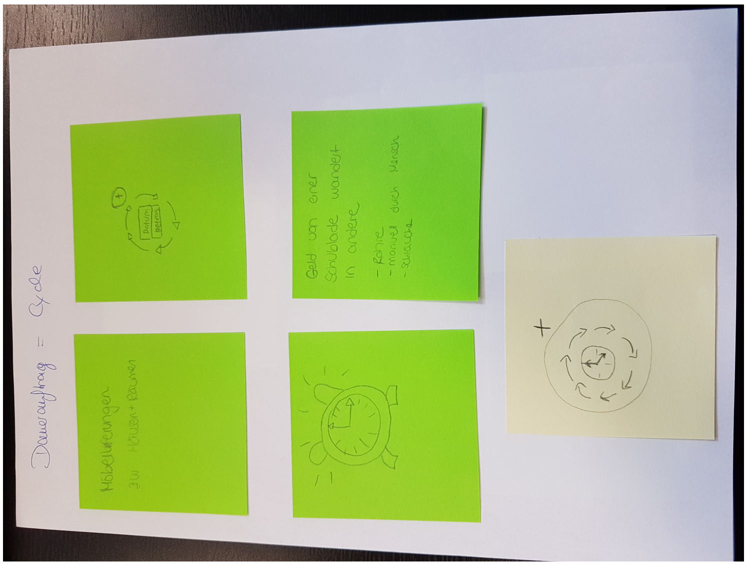
Intuitive Benutzung	sehr schlecht	□ (1)	□ (2)	□ (3)	□ (4)	□ (5)	□ (6)	□ (7)	sehr gut
Innovation	sehr schlecht	□ (1)	□ (2)	□ (3)	□ (4)	□ (5)	□ (6)	□ (7)	sehr gut
Bedienbar durch ältere Menschen (>65)	sehr schlecht	□ (1)	□ (2)	□ (3)	□ (4)	□ (5)	□ (6)	□ (7)	sehr gut

Vielen Dank!

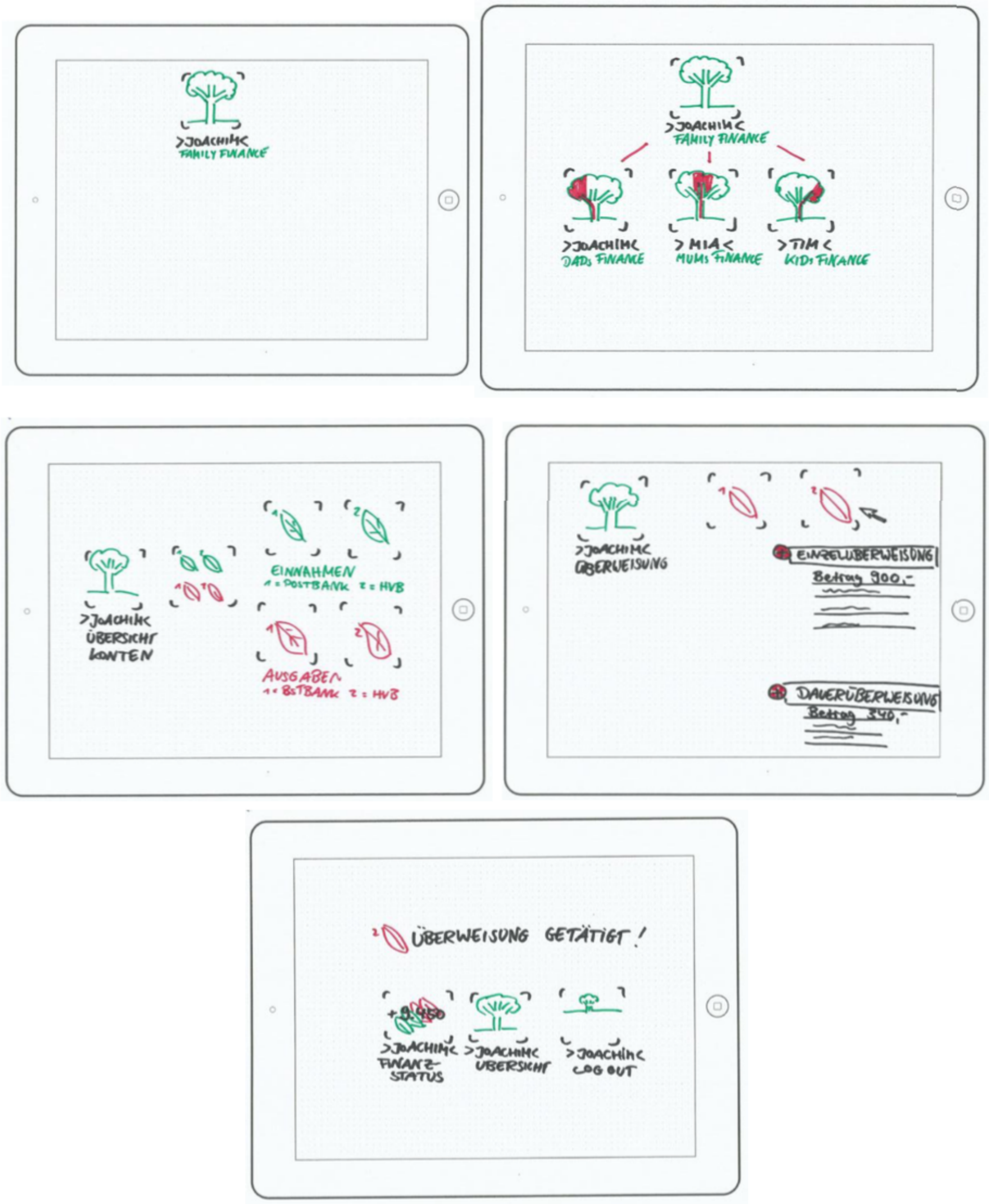
2

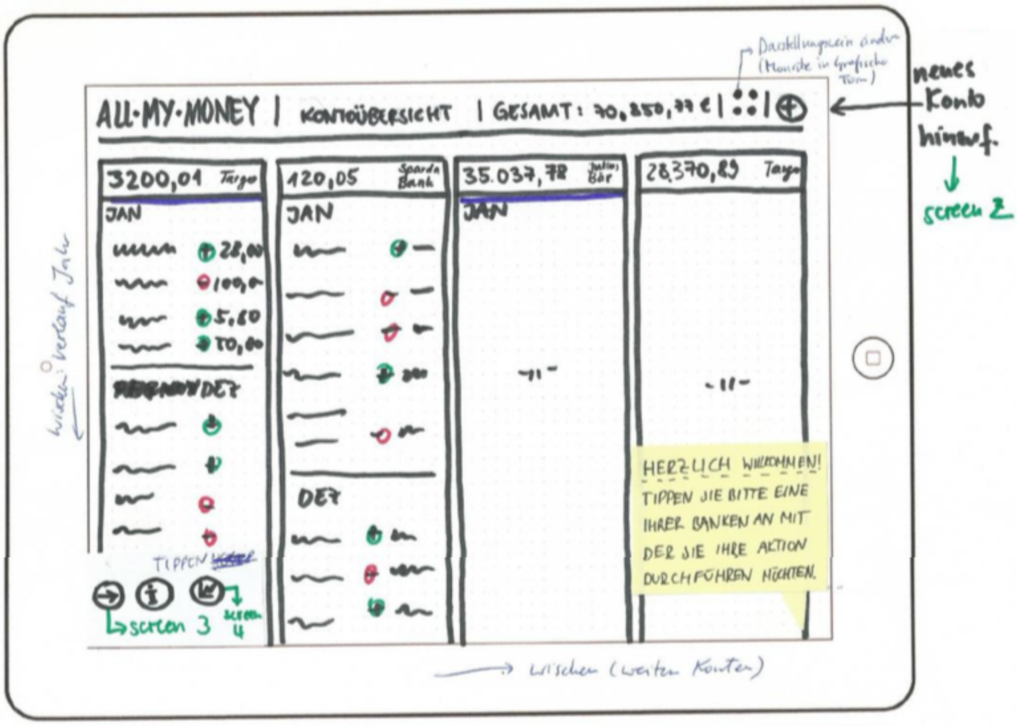
A.3.5 Personas

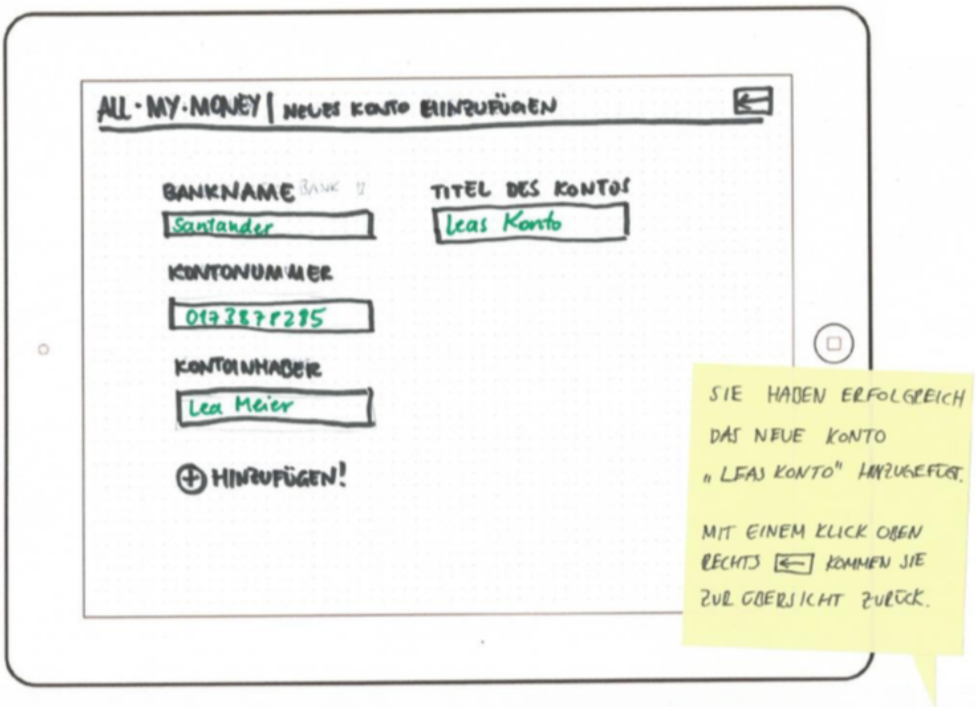
A.3.6 Exemplary Design Ideas during the Wall Walk



A.3.7 Exemplary Prototypes




ALL·MY·MONEY
KONTOÜBERSICHT
GESAMT:
3200,01
120,05
35.037,78
28,370,89
JAN
DEZ
neues Konto hinzuf.
screen 2
HERZLICH WILLKOMMEN!
TIPPEN SIE BITTE EINE IHRER BANKEN AN MIT DER SIE IHRE AKTION DURCHFÜHREN MÖCHTEN.
screen 3
wischen (weitere Konten)


ALL·MY·MONEY | NEUES KONTO EINZUFÜGEN
BANKNAME
Santander
TITEL DES KONTOS
Leas Konto
KONTONUMMER
KONTOINHABER
Lea Meier
HINZUFÜGEN!
SIE HABEN ERFOLGREICH DAS NEUE KONTO „LEAS KONTO" HINZUGEFÜGT.
MIT EINEM KLICK OBEN RECHTS KOMMEN SIE ZUR ÜBERSICHT ZURÜCK.

A.3.8 Questionnaire specific for conditions with image-schematic metaphors used after the study

VP-Code:________________________ (**Vo**rnameMutter, **Vo**rnameVater, **Geburtsmonat**: z.B. MaPe05)

Datum:__________________ Team:_______ Bedingung:__________________

Nachbefragung – Image Schema Methode

In diesem Fragebogen sollen Sie die gerade eben verwendete Image Schema Methode bewerten.
Bitte machen Sie zu jeder Frage eine Aussage.

Wie gut kannten Sie die Image Schema Methode bereits?

gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr gut

Was war besonders gut an dieser Methode?

Was war besonders schlecht an dieser Methode?

Was sollte an dieser Methode verbessert werden?

Was war besonders gut/schlecht an der selbstständigen Prototypenerstellung?

Was war besonders gut/schlecht an der gemeinsamen Evaluation der Prototypen?

Wie sehr hat diese Methode den Arbeitsprozess strukturiert?

Gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr stark

Wie sehr hat diese Methode die Kreativität angeregt und Sie auf neue Gedanken gebracht?

Gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr stark

VP-Code:________________________ (**Vo**rnameMutter, **Vo**rnameVater, **Geburtsmonat**: z.B. MaPe05)

Datum:__________________ Team:_______ Bedingung:__________________

Wie hilfreich war diese Methode insgesamt?

Gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr stark

Wie leicht verständlich war diese Methode?

Gar nicht □ (1) □ (2) □ (3) □ (4) □ (5) □ (6) □ (7) sehr stark

Was könnte man an der Instruktion verbessern?

Vielen Dank!

A.4 Appendix Study 5

No Image-Schematic Metaphors provided *Image-Schematic Metaphors provided*

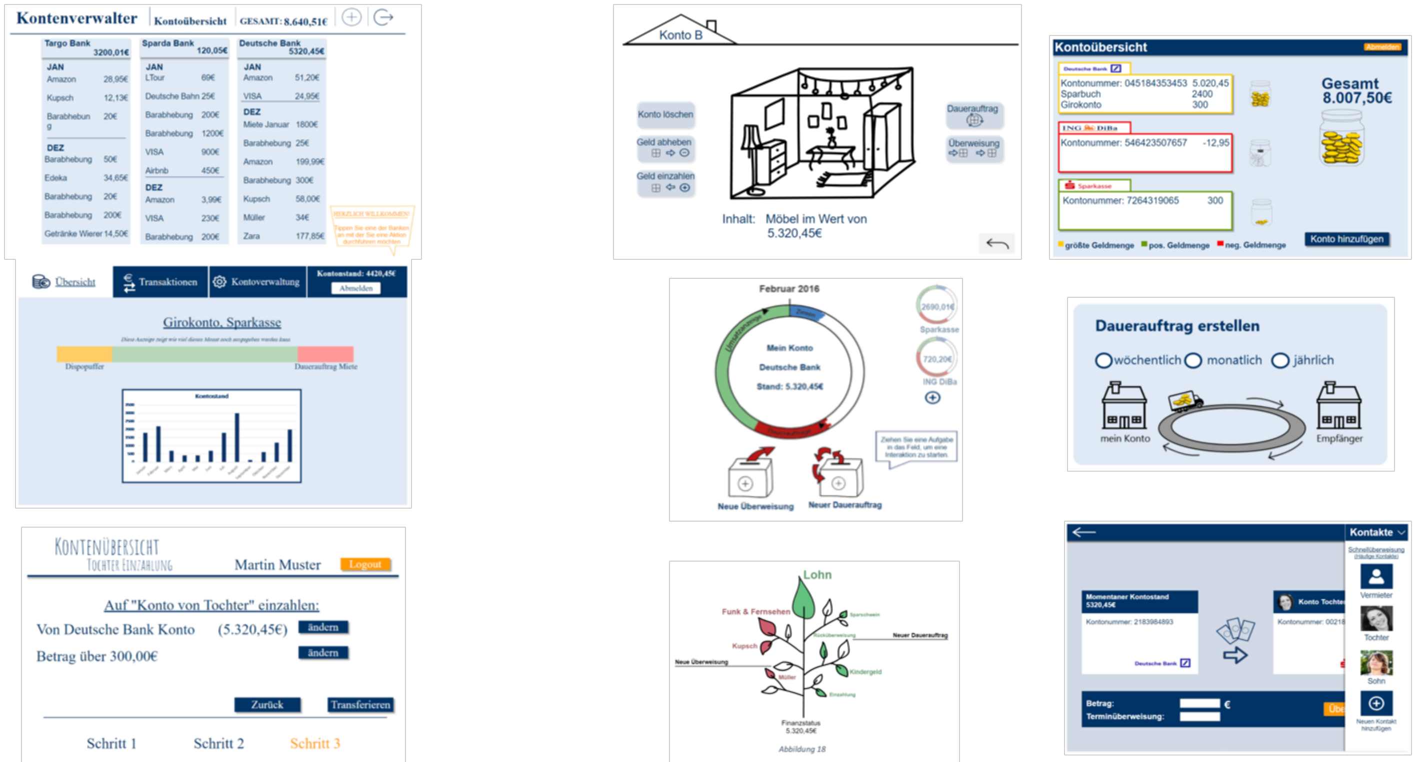


Figure A.1: Prototypes that had been designed explicitly with or without image-schematic metaphors used in study 5. All prototypes were evaluated on an iPad.

Table A.1: Descriptive data (time, mental workload and perceived innovation) per design condition, prototype and age group (mean and standard deviation). SMEQ: Subjective Mental Effort Questionnaire, HQS: subscale "hedonic quality - stimulation" of the AttrakDiff2-questionnaire (perceived innovation). PT: Prototype. IS-M: Image-schematic Metaphor.

		Time [s]		SMEQ [0-220]		HQS [1-7]	
		Younger Adults	Older Adults	Younger Adults	Older Adults	Younger Adults	Older Adults
Affinity Diagram Only	PT1	158.1 [25.7]	238.8 [39.8]	49.8 [32.4]	35.2 [16.8]	3.5 [1.0]	4.4 [0.9]
	PT2	164.6 [30.5]	297.5 [81.6]	40.3 [29.5]	36.8 [18.2]	3.2 [0.9]	4.4 [1.2]
	PT3	189.7 [54.7]	354.4 [117.0]	68.8 [43.3]	65.5 [37.0]	3.8 [0.9]	4.6 [1.0]
	PT4	185.5 [31.7]	348.7 [110.9]	34.6 [19.1]	33.3 [12.5]	3.0 [1.0]	3.9 [1.4]
Affinity Diagram & IS-M	PT1	174.8 [35.1]	287.7 [80.7]	57.8 [30.5]	54.8 [50.5]	4.4 [1.1]	5.1 [1.2]
	PT2	161.8 [26.1]	277.3 [65.3]	66.0 [35.6]	38.8 [19.9]	5.2 [0.8]	5.1 [1.2]
	PT3	203.3 [56.4]	250.5 [70.4]	62.8 [25.5]	39.4 [22.1]	3.5 [1.0]	4.4 [0.7]
	PT4	228.5 [51.3]	406.1 [96.7]	61.2 [42.3]	50.1 [17.8]	5.5 [0.7]	5.5 [0.6]
Only IS-M	PT1	141.2 [34.1]	209.6 [46.7]	68.2 [40.0]	86.3 [50.9]	5.2 [0.6]	5.2 [0.8]
	PT2	134.8 [22.1]	237.6 [58.6]	54.8 [33.0]	47.4 [35.1]	5.4 [0.7]	5.4 [0.9]
	PT3	140.4 [20.3]	283.0 [104.8]	64.8 [31.9]	50.7 [20.4]	5.1 [0.7]	5.0 [1.0]
	PT4	186.8 [49.8]	277.0 [48.5]	53.7 [36.8]	28.2 [11.1]	3.8 [0.8]	4.7 [1.3]

A.5 Appendix Study 6

A.5.1 Use case instruction

Use Case 1 You are going on vacation. For routing you use your navigation system. While on the road, you get hungry. Your navigation system shows you recommended restaurants along the route. Choose a restaurant and let yourself be navigated there.

Use Case 2 You are listening to music while driving. You want to listen to the track that was played second to last. Choose the second to last track and play it.

Use Case 3 You hear a song you like very much and would like to show it to friends. Share the song currently playing with the four persons on the list you received from the experimenter.

Use Case 4 You get in to a congestion on the route. However, it is not displayed by your navigation system. Report the congestion on your route.

Use Case 5 You would like to call several persons in order to talk something over with them. Call Dave first. Well done, now add Laura [and then] Tom to the call. Well done, now remove Dave from the call.

Instructions were identical for prototypes of both design conditions.

A.6 Appendix Synthesis Revisited

This section aims at providing concrete and brief guidance for practitioners that want to apply the method of image-schematic metaphors in their projects that are in line with User-Centered Design (e.g. Contextual Design Holtzblatt & Beyer, 2017). The focus will be on extracting image-schematic metaphors from natural language and not on observations or documented corpora.

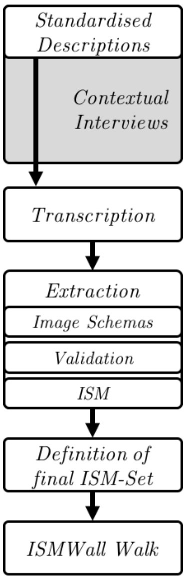


Figure A.2: Steps for extracting and applying image-schematic metaphors in HCI projects.

The recommendations and schedules of this section are derived from published and unpublished documentations of a variety of projects (Hurtienne & Blessing, 2007; Hurtienne, Klöckner et al., 2015; Klöckner, Maier & Nass, 2010; Löffler, Hess, Hurtienne et al., 2013; Tscharn, 2017; Winkler et al., 2016) and represent the best practices for applying the method of image-schematic metaphors. The exact procedure will depend on the specific project. The following steps are intended to provide feasible guidance primarily for HCI-related projects. These steps do not claim to be as rigorous as basic linguistic methodologies (which are not feasible in practice) like the MIPVU procedure described in chapter 4 (Steen et al., 2010).

The method should be blended into two phases of User-Centered Design: a) user research and b) ideation. In the user research phase, user utterances are analysed for domain specific and project relevant image-schematic metaphors. In the ideation and prototyping phase, these metaphors serve as the basis for stimulating creativity while at the same time constraining design ideas to those that are in line with the user's universal mental representation of abstract concepts. This two high-level phases will be translated into specific step-by-step directives in the following with the intention to facilitate the adoption of the method of image-schematic metaphors in the future. The single steps will be:

1. Contextual interviews and standardised descriptions
2. Transcription
3. Extraction of image-schematic metaphors
4. Defining the final set of image-schematic metaphors
5. Ideation: Image-schematic metaphors Wall Walk

A.6.1 Contextual interviews

Ideally, 4-12 participants of different age groups are available for the user research phase (see further, Holtzblatt & Beyer, 2017). The interviews must be audio-recorded. When core concepts of the later interface (e.g. LOUDNESS, FRIENDSHIP) are already known before contextual interviews, the interviewer should ask the participants to briefly describe those using their own words ("Please, explain to me using your own words what an [abstract concept] is"). The interviewer should follow the usual procedure of the contextual interview (Holtzblatt & Beyer, 2017).

Example: Interviews on communication between different cars with ten participants. Standardised description of the abstract group conversation "Please explain to me, in your own words: what is a GROUP CONVERSATION?".

A.6.2 Transcription

The interviews must be transcribed on a word-to-word basis. This means that utterances such as "uhm" or "well" (or their corresponding utterances in other languages) are left out, but otherwise, no change of wording should occur during transcription. Conserving the exact wording of the user utterances is important. When descriptions of the core concepts are available, they should be transcribed and analysed first. Contextual interviews should not be transcribed completely, before the main functionality of the system (and, thus, the abstract concepts of interest) are defined, and the standardised descriptions do not cover core abstract concepts. Parts that do not cover those abstract concepts can be left out for transcription and image-schematic metaphor analysis.

Example: In the transcription of the standardised definitions, one participant stated: "Group conversation is when at least three people talk with each other. This can happen face-to-face, but also over the distance. Also... sometimes it is spontaneous to talk with other people. For example, in a group discussion in your company via Skype, you can always ask an external expert into the discussion circle and add them to the conversation.". The recording must be transcribed on a word-to-word level.

A.6.3 Extraction of image-schematic metaphors

Setup Ideally, the persons responsible for extraction of image-schematic metaphors should be trained to gain a basic understanding of image schemas and image-schematic metaphors. However, this will often be not feasible in practice. When this is the case, the image schema definitions of Hurtienne (2011) can provide a good start to understand the essence of image schemas and image-schematic metaphors. For training purposes, inexperienced team members should conduct and participate in short interviews within the team and analyse the transcript of their natural language. This can increase their consciousness about both their language and behaviour during the contextual interviews as well as serves as a first encounter with the extraction method before analysing the standardised descriptions and contextual interviews of the current project.

For the extraction process itself, one person can suffice but, especially for inexperienced teams, two or three extractors allow discussions about both ambiguities in the extraction process itself as well as the final set of extracted image-schematic metaphors. The extraction process should start with the standardised descriptions. In later phases of the project, contextual interviews can also be analysed, but extraction should not start before the core abstract concepts are defined. The extraction procedure is identical for standardised descriptions of abstract concepts or relevant sections in transcripts and consists of three iterations.

Identification of image schemas First, the descriptions or relevant parts of the transcripts should be analysed for only image schemas and not directly for image-schematic metaphors. Every word, pre- and suffix should be analysed for image-schemas. Note that not all words allow for the extraction of image schemas. For example, "With GitHub, I can push code to the repository" does not contain image schemas in words "GitHub", "I", " code", "the" and "repository". Still, "with" might refer to LINKAGE, "can" to ENABLEMENT, "push" to

COMPULSION and "to" to PATH. Depending on what the abstract concepts of interest are, the respective image schema can be later used to form an image-schematic metaphor. Compare this sentence to the slightly alternated form: "With GitHub, I can push code *in*to the repository". Here, "into" can be split into its syllables "in-" and "-to". This adds the word "into" the image schema IN-OUT as well as "repository" the image schema CONTAINER and PATH.

Importantly, only metaphorical expressions should be annotated with image schemas. For example, "I pour water *into* a cup" contains the keyword *in* for the image schemas IN-OUT and CONTAINER, but is not metaphorical but a physical process. Here, no image schema should be annotated because it cannot lead to an image-schematic metaphor later. In contrast, "This feature was introduced in 2017" contains the same keyword *in*, but it does not describe a physical fact but is metaphorical. Thus, in this case, the image schemas IN and CONTAINER are appropriate to extract from this expression. In the previous example "With GitHub, I can push code to the repository", no physical object is pushed in the physical world so that this expression can be regarded as metaphorical. When it is unclear whether an expression is metaphorical, dictionaries can help to clarify (e.g. Duden, 2017).

Especially image schemas from the FORCE class often require considering also the context of the word. For example, in the expression "As soon as I enter the Captcha code, I can go on with the registration process", the extractor can find the image schema RESTRAINT REMOVAL. A BLOCKAGE (in this case the Captcha) usually precedes RESTRAINT REMOVAL. However, without taking into account the context, no single keyword can be defined in this case that would always justify the extraction.

Example: An annotated sentence could look like this: "in [CONTAINER] a group discussion [CONTAINER] in [IN-OUT] your company [CONTAINER] via [PATH] Skype, you can [ENABLEMENT] always ask an external [IN-OUT] expert into [IN-OUT] the discussion circle [CONTAINER] and add [MERGE] them to [PATH] the conversation. You can [ENABLEMENT] just push [COMPULSION] them into [IN-OUT] the conversation [CONTAINER]". For standardised descriptions, only image schemas with a connection to the abstract concept are relevant (here, GROUP CONVERSATION) and the others can be left out during the extraction process (e.g. the image schema CONTAINER for "company" is not connected to group conversation and could be ignored during extraction).

Validation of image schemas As soon as the extractor has analysed all relevant parts of the transcript for image schemas, a second iteration can increase internal consistency and the reliability of the annotations. This second iteration is highly recommendable when resources are available to increase the reliability of the extracted image schemas. Two procedures can be recommended: first, each annotation should be rechecked whether the word or phrase is metaphorical. Second, the extractor should check whether the same phrase always led to the extraction of the same image schema. Sometimes an extractor finds reoccurring phrases and does no extract image schemas for all of them. Since the frequency of image-schematic metaphors helps in prioritising for the final set of image-schematic metaphors, the extractor should ensure a standardised and consistent extraction process, especially, when only one extractor is involved. Note that this validation step further increases the needed resources for the extraction process and can be omitted when resources are already scarce and limited. However, these two quality measures can help to increase the reliability and consistency of the extraction process.

Example: After completing extraction for all descriptions, the annotated texts are re-read to check whether all extracted image schemas are based on metaphorical expressions and the same phrases always led to the same image schemas.

Formulating image-schematic metaphors For each annotation of an image schema, the extractor formulates an image-schematic metaphor. For this step, it can be helpful that relevant target domains are already defined. A clear focus on relevant target domains can avoid metaphors that are irrelevant for the later interaction design. For example, one core functionality of an online-banking app might be displaying the current interests the users get on top of their money. When the extractor is aware of this focus on the target domain INTERESTS, the extractor can avoid image-schematic metaphors like MONEY IS ENABLEMENT (e.g. from "I *can* buy things with money) and extract only those metaphors that have the form of INTEREST IS ... (e.g. from "*Every month*, I get 4% interest *on top* of my money in the bank account: INTEREST IS CYCLE, INTEREST IS UP). Still, the extractor should check every image schema whether she can extract an appropriate image-schematic metaphor from it.

Example: Starting with the sentence described above (description of GROUP CONVERSATION), several image-schematic metaphors could be defined for the abstract concept of group communication: GROUP CONVERSATION IS CONTAINER, PARTICIPANT OF GROUP CONVERSATION IS IN, ADDING PARTICIPANT TO GROUP CONVERSATION IS COMPULSION ON A PATH and GROUP CONVERSATION IS ENABLEMENT.

A.6.4 Defining the final set of image-schematic metaphors

Finally, the single annotations of image-schematic metaphors must be distilled into a set of image-schematic metaphors relevant for later interaction design. Because the transcripts will offer many image-schematic metaphors with different frequencies, the project team can draw on several rules for prioritising.

1. Focus on image-schematic metaphors that address a highly relevant abstract concept of your project.
2. Focus on image-schematic metaphors with a high number of instantiations.
3. Focus on image-schematic metaphors that might be interesting and stimulate visual thinking.

The output of the first phase of the image-schematic metaphor method is the final list of image-schematic metaphors. To illustrate each image-schematic metaphor, a user utterance that shows its origin can be useful.

Example: For the example sentence, all metaphors focus on the abstract concept of GROUP CONVERSATION. Since GROUP CONVERSATION IS CONTAINER and PARTICIPANT OF A GROUP CONVERSATION IS IN occur multiple times and is directly relevant to the design process, these metaphors should be included in the design process. The metaphor ADDING A PARTICIANT TO GROUP CONVERSATION IS COMPULSION ON A PATH is less frequent but could be highly relevant for a function like adding a participant. GROUP CONVERSATION IS ENABLEMENT is also frequent but the benefit for interaction design might be not as high as the other two metaphors. For this example, the metaphors GROUP CONVERSATION IS CONTAINER, PARTICIPANT OF A GROUP CONVERSATION IS IN and ADDING A PARTICIPANT TO GROUP CONVERSATION IS COMPULSION ON A PATH will be prioritised.

A.6.5 Ideation with image-schematic metaphors

As soon as the final list of image-schematic metaphors is defined, it can be used for interaction design. Depending on the main method, the specific integration into the design process will differ. However, based on this work, the recommendation is to plan a separate activity for designing with image-schematic metaphors. By this, the potential of the method will be maximised. The following procedure gives an example for such an activity, but the particular procedure can deviate from this example and might depend on the concrete project and context. The core functionality and requirements for the user interface should be defined (e.g. by other activities of the User-Centred Design process).

Interaction designers should be briefly introduced to the theory of image schemas and image-schematic metaphors. Besides introducing them to the topic, this will increase the seriousness they will show when translating single image-schematic metaphors into design ideas.

As described above, a single activity focusing on image-schematic metaphors might show better results than the integration of image-schematic metaphors into an existing method for interaction design (such as an affinity diagram wall walk). Instead, an image-schematic metaphors wall walk can provide the structure necessary for coming up with a variety of design ideas that instantiate each image-schematic metaphor in the user interface. The number of design ideas turns out to be most essential and enough time should be provided to complete the image-schematic wall walk. For the wall walk, each image-schematic metaphor and one exemplary expression illustrating the metaphor should be printed on a sheet of paper and attached to the wall with enough surrounding space for design ideas that can also be attached to the wall. After the Wall Walk, team members should prioritise the single ideas.

Example: Three members of the project team participate in the Wall Walk. For the visualisation of the members of a group conversation, they generate design ideas according to the metaphors GROUP CONVERSATION IS CONTAINER, PARTICIPANT OF A GROUP CONVERSATION IS IN and ADDING A PARTICIPANT TO GROUP CONVERSATION IS COMPULSION ON A PATH. After the Wall Walk, each team member receives ten coloured dots to distribute overall most exciting design ideas. Most dots are placed on the design idea that combined all three metaphors. This design idea is further developed and leads to figure A.3.

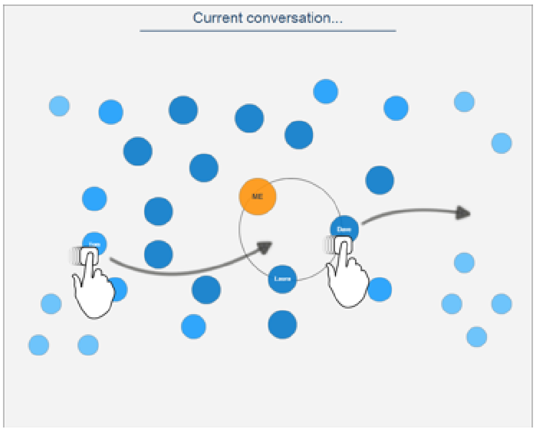


Figure A.3: Visualised design idea that combines all three metaphors from the example. GROUP CONVERSATION IS CONTAINER *is instantiated via the circle with an in and outside.* PARTICIPANT OF A GROUP CONVERSATION IS IN *is instantiated via the objects (single participants) that can be in or outside of the container.* ADDING A PARTICIPANT TO GROUP CONVERSATION IS COMPULSION ON A PATH *is instantiated via the drag-and-drop gesture that allows to move participants into the circle. Another important image-schematic metaphor (*GROUP CONVERSATION IS LINKAGE*, based on expressions like "we talk with each other") is also instantiated in this design idea.*

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Würzburg, den 17. April 2018

Robert Tscharn