REQUIREMENTS FIXATION: THE EFFECT OF SPECIFICATION FORMALITY ON DESIGN CREATIVITY

Rahul Prem Mohanani
RAHUL PREM MOHANANI

REQUIREMENTS FIXATION: 
THE EFFECT OF SPECIFICATION 
FORMALITY ON DESIGN 
CREATIVITY

Academic dissertation to be presented with the assent of 
the Doctoral Training Committee of Information 
Technology and Electrical Engineering of the University of 
Oulu for public defence in Martti Ahtisaari -sali (L2), 
Linnanmaa, on 9 December 2019, at 12 noon

UNIVERSITY OF OULU, OULU 2019
Mohanani, Rahul Prem, Requirements fixation: The effect of specification formality on design creativity.
University of Oulu Graduate School; University of Oulu, Faculty of Information Technology and Electrical Engineering
Acta Univ. Oul. A 734, 2019
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

There is a broad consensus in the software engineering (SE) research community that understanding system desiderata and design creativity is critical for the success of software projects. This has motivated a plethora of research in SE to improve requirements engineering (RE) processes. However, little research has investigated the relationship between the way desiderata are presented (i.e., framed) and creative design performance. This dissertation, therefore, examines the effects of more formal presentations of desiderata on design creativity.

The research was conducted in three phases. The first consisted of summarizing the available literature on cognitive biases in SE to build a comprehensive body of knowledge, understand the current state of research, and identify the relevant literature to position and delineate subsequent investigations involving the framing effect and fixation. This research phase also investigated how creativity is conceptualized (i.e., understood, assessed and improved) in SE by exploring the perceptual differences and similarities between SE researchers and practitioners. In the second phase, two controlled experiments were conducted to investigate the impact of framing desiderata first as requirements (in general) and then as prioritized requirements on design creativity (i.e., the originality and practicality of a design). The third phase involved a protocol study to explore the underlying cognitive mechanisms that may explain why framing desiderata as formal requirements affects creativity. The empirical evidence from the second and third phases was interpreted together to propose a theoretical framework that explains the effect of specification formality on design creativity.

While the results of the experiments show that specification formality is negatively related to design creativity (i.e., desiderata framed as requirements or prioritized requirements result in designs that are less creative), the findings from the protocol study indicate that the negative relationship between specification formality and design creativity is mediated by fixation (i.e., more formal presentation of desiderata induces fixation and hinders critical thinking). Overall, the results of this dissertation suggest that more formal and structured presentations of desiderata cause requirements fixation—the tendency to attribute undue confidence and importance to desiderata presented as formal requirements statements—that affects design creativity, and thus undermines software engineering success.

Keywords: cognitive bias, creativity, experiment, fixation, framing effect, interview, originality, practicality, prioritization, protocol study, requirements, software design, software engineering, systematic mapping
Mohanani, Rahul Prem, Vaatimusten fiksaatio: Määrittelyjen muodollisuuden vaikutus suunnittelun luovuuteen.
Oulun yliopiston tutkijakoulu; Oulun yliopisto, Tieto- ja sähköteknikan tiedekunta
Acta Univ. Oul. A 734, 2019
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Ohjelmistotuotannon tutkijoiden keskuudessa on laaja yksimielisyys järjestelmän tarpeiden ja suunnittelun luovuuden ymmärtämisen kriittisydestä ohjelmistoprojektien menestyksessä. Tämä on motivoineet monia ohjelmistotuotannon vaatimustenmäärityspyynnösten parantamiseen liittyviä tutkimuksia. Harvassa on tarkasteltu tarpeiden esitystavan (eli muotoilun) ja luovan suunnittelun lopputuloksen välistä yhteyttä. Tässä väitöskirjassa tarkastellaan tarpeiden muodollisempien esitystapojen vaikutuksia suunnittelun luovuuteen.

Tutkimus oli kolmivaiheinen. Ensinnä referoitiin ohjelmistotuotannossa kognitiivisiin harhoihin liittyvä kirjallisuus kartoittamaan nykytutkimuksen tila ja merkityksellinen kirjallisuus myöhempien, kehysvaikutusten ja fiksaatiota sisältävien tutkimusten järjestelyseen ja rajaamiseen. Lisäksi tarkasteltiin luovuuden käsittelystä (elä ymmärrävyttä, arviointia ja parantamista) tutkimalla katsannollisia eroja ja yhtäläisyyksiä tutkijoiden ja ammattilaisen välillä. Toisessa vaiheessa tehtiin kaksi kontrolloitua koetta tarpeiden muotoilun vaikutuksen tutkimiseksi, ensin vaatimukseina (yleisesti) ja sitten tärkeysjärjestykseen laitettuina vaatimuksina suhteessa suunnittelun luovuuteen (eli omaperäisyyteen ja käytännöllisyyteen). Lopuksi, protokollatutkimuksella selvitettiin taustalla olevia kognitiivisia mekanismeja selittämään syitä muodollisina vaatimuksina esitettyjen tarpeiden vaikutuksista luovuuteen. Toisesta ja kolmantesta vaiheesta saatuja empiristen aineistojen tulkittiu yhdessä muodostavan teoreettisen viitekehyksen, joka selittää määrittelyyn muodollisuuden vaikutuksista suunnittelun luovuuteen.

Vaikka kokeiden tulokset osoittavat määrittelyjen muodollisuuden vaikuttavan negatiivisesti suunnittelun luovuuteen (eli tarpeiden muotoilua vaatimuksina tai priorisoituina vaatimuksina vähentää suunnittelun luovuutta), protokollatutkimuksen tulokset viittaavat fiksaation vaikutuksiin negatiiviseen suunnittelun määrittelyjen muodollisuuden ja suunnittelun luovuuden välillä (eli tarpeiden muodollisempi esitystapa aiheuttaa fiksaatiota ja vaikeuttaa kriittistä ajattelua). Kuitenkin, vaatimustenmäärityksen tulokset aiheuttavat ja vaikuttevat määrittelyjen ja fiksaation välille

Asiakas: fiksaatio, haastattelu, hehkysvaikutus, koe, kognitiivinen harha, käytännöllisyys, luovuus, ohjelmistosuunnittelu, ohjelmistotuotanto, omaperäisyys, priorisointi, protokollatutkimus, systemaattinen kartoitus, vaatimukset
There is no companion equivalent to knowledge, there is no friend equivalent to knowledge. There is no wealth equivalent to knowledge, there is no happiness equivalent to knowledge

—Unknown
This Thesis is Dedicated to my Parents
Acknowledgements

“But I know you can...I believe in your (research) ideas”, was the reply by Prof. Burak Turhan, back in 2014, when I asked him about the reason he selected me as his supervisee among many other potential applicants. These words alone motivated and pushed me through the next five years to complete my PhD degree. Completing my PhD involved a lot more than simply conducting research and disseminating the results. I consider the last five years as a perfect sample-set of life, which not only gave me the opportunity to grow both intellectually and personally but also made me overcome my fears, self-doubts and weaknesses. I would like to express my sincere gratitude to all the people who supported me during this journey.

First, I am grateful to my principal supervisor, Prof. Burak Turhan for his complete faith in my research and his unconditional support during all stages of my doctoral journey. This dissertation is a direct result of his valuable guidance and consistent encouragement. Thank you, Prof. Turhan, for your patience during our numerous discussions, my silly questions, your understanding of my shortcomings and your endless motivation. Along with Prof. Turhan, I owe this dissertation to my co-supervisor, Dr. Paul Ralph. Much of this research work would not have been possible without Dr. Ralph’s scholarly inputs and his faith in me and my research skills. Thank you Dr. Ralph, for your trust, the continuous exchange of ideas, the successes, the criticisms and the failures. You been there for me whenever I have needed your guidance. I would also like to convey my sincere gratitude to my co-supervisor, Dr. Pilar Rodríguez, for her constant help to clarify my doubts, continuous feedback, and polishing my research ideas.

I would like to extend a special thank you to Prof. Markku Oivo for accepting me as a member of one of the leading software engineering research groups. It was a privilege to be a part of the M3S group. I would also like to thank my follow-up group members: Prof. Mika Mäntylä and Prof. Minna Isomursu. Our annual meetings and your feedback served as the most valuable gauge of my progress, which steered my research in the right direction. Thank you very much for your support and encouragement. I am sincerely grateful to Prof. Günther Ruhe (University of Calgary, Canada) and Prof. Marco Kuhrmann (University of Passau, Germany), both of whom dedicated their valuable time to review this dissertation. Your detailed comments and suggestions
allowed me to further improve the quality of this dissertation. My sincere thanks goes to Professor Pekka Abrahamsson (University of Jyväskylä, Finland) for agreeing to act as an opponent of this dissertation. Additionally, I would like to thank my fellow researchers who have co-authored some of the publications with me. I acknowledge Iflaah Salman, Prabhat Ram and Ahmed Lasisi (University of Oulu, Finland), Dr. Ben Shreeve (Lancaster University, United Kingdom), and Dr. Vladimir Mandić (University of Novi Sad, Serbia), who have sincerely contributed in the various publications present in this dissertation. I would also like to acknowledge Oulu Infotech Doctoral Program for funding my doctoral research, as well as University of Oulu Graduate School (UniOGS) travel grant and HPY: Research Foundation Grants (HPY:n Tutkimussäätiö Apurahat) that supported my research visit to the University of Auckland, New Zealand.

This journey would not have been possible without the love and support of my friends at the University of Oulu. I would forever cherish that every cup of coffee, every morsel of food, and every bit of laughter and sorrows that we shared with each other. A huge thanks to (in no specific order) Adrian, Ivana, Alireza, Ovais, Lucy, Vitor, Woubshet, Pertti, Iflaah, Prabhat, Ahmad, Rebvar, Itir, Tanja, Elina, Päivi, Anna, Hanna-Leena and many others. I wish all the best for each of you. Special thanks are in order for Dr. Davide Fucci, Dr. Nebojsa Tausan and Sandun Dasanayake. Thank you for hearing me out and for being there. You are family. I would also like to thank all of my other colleagues and senior members from the M3S research group.

My journey as a part of the University of Oulu was interwoven with many visits to my hometown—Mumbai, India. For this reason, I would like to thank my lifelong friends: Dheeraj Punjabi, Poonam Punjabi, Rohan Mali, Vinay Bajaj, Harshita Bajaj, Sairus Joni, Rahul Chawla and Udit Varma, for making me feel as if I have never left.

Finally, my warmest appreciation goes to my family. First, to the love of my life, my wife, Shweta Mohanani. I know its not been easy for you; however you always stood by me. Thank you for being an inexhaustible source of love and my strongest ally in this journey. Lastly, immense thanks to my parents, who have always supported my dreams and ambitions unconditionally. All my success and happiness are because of your blessings and love. I know that me living so far away from home has been tough on you. That is why I dedicate this dissertation to you both. I love you a lot. Thank you for everything.

Rahul Mohanani
New Delhi, India, October 2019
### List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>DV</td>
<td>Dependent Variable</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-Computer Interaction</td>
</tr>
<tr>
<td>IV</td>
<td>Independent Variable</td>
</tr>
<tr>
<td>IS</td>
<td>Information Systems</td>
</tr>
<tr>
<td>NHST</td>
<td>Null Hypothesis Significance Testing</td>
</tr>
<tr>
<td>RE</td>
<td>Requirements Engineering</td>
</tr>
<tr>
<td>SMS</td>
<td>Systematic Mapping Study</td>
</tr>
<tr>
<td>SE</td>
<td>Software Engineering</td>
</tr>
<tr>
<td>SWEBOK</td>
<td>Software Engineering Book of Knowledge</td>
</tr>
<tr>
<td>SLR</td>
<td>Systematic Literature Review</td>
</tr>
</tbody>
</table>
List of original publications

This dissertation is based on the following articles, which are referred to in the text by their Roman numerals (I–V):


## Contents

Abstract  
Tiivistelmä  
Acknowledgements  
List of abbreviations  
List of original publications  
Contents  
1 Introduction  
1.1 Research gaps  
1.2 Research objective and research questions  
1.3 Scope of the research  
1.4 Dissertation structure  
2 Background and related work  
2.1 Cognitive biases  
2.1.1 Framing effect  
2.1.2 Fixation  
2.2 Task structuring  
2.3 Design thinking  
2.4 Creativity  
3 Research design  
3.1 Research phases  
3.1.1 Phase 1: Identifying the relevant literature and variables; Papers I and II  
3.1.2 Phase 2: Experimentation; Papers III and IV  
3.1.3 Phase 3: Protocol study; Paper V  
3.2 Research process  
3.2.1 Systematic mapping study (SMS)  
3.2.2 Cross-sectional interviews  
3.2.3 Controlled experiments  
3.2.4 Protocol study  
3.3 Summary of the research process
4 Original research publications

4.1 Author’s contributions

4.2 Paper I - Cognitive biases in software engineering: A systematic mapping study

4.3 Paper II - Perceptions of creativity in software engineering research and practice

4.4 Paper III - Requirements fixation

4.5 Paper IV - Requirements framing affects design creativity

4.6 Paper V - Requirements fixation: A dialog-based protocol study

5 Results, discussions, and limitations

5.1 Answer to RQ1: What is the current state of research concerning cognitive biases and creativity in SE?

5.1.1 Answer to RQ1.1: What is the current state of the art in SE research involving cognitive biases?

5.1.2 Answer to RQ1.2: How is creativity conceptualized, assessed, and improved in SE?

5.2 Answer to RQ2: How does specification formality affect design creativity?

5.2.1 Answer to RQ2.1: Does the framing of desiderata affect design creativity?

5.2.2 Answer to RQ2.2: What cognitive behaviours explain the negative relationship between specification formality and design creativity?

5.3 Summary of results

5.4 Threats to validity

5.4.1 Systematic mapping study

5.4.2 Interviews

5.4.3 Experimentation

5.4.4 Protocol study

6 Conclusion and future research

6.1 Summary of contributions

6.1.1 Implications for SE education

6.1.2 Implications for SE practice

6.2 Future research avenues

References
Appendices  
Original publications
1 Introduction

The software engineering (SE) research community widely acknowledges that understanding system requirements is critical and essential for designing good software systems (Chow & Cao, 2008). The notions that requirements engineering (RE) must produce “unambiguous”, “consistent”, “feasible”, “complete”, “traceable”, and “verifiable” requirements specifications (IEEE, 1998, p. 1), and that clear and structured requirements lead to successful software systems (Mund, Fernandez, Femmer, & Eckhardt, 2015) are strongly entrenched in SE research. This belief has stimulated a variety of research concentrated on improving RE processes. One of the more successful attempts to clarify and structure the requirements, especially in software-intensive systems, is by prioritizing requirements based on their perceived importance (Herrmann & Daneva, 2008). In other words, RE research has continually attempted to develop better methods, techniques, and approaches for eliciting, analysing, modelling, and communicating requirements that are clear, structured, unambiguous, and definitive. Similarly, practitioners consider understanding the requirements earlier in the process as critical to a system’s success, especially when outsourcing and tendering contracts (Brooks Jr, 2010). Moreover, the dangers of getting the requirements wrong or failing to accommodate changes to requirements as needs evolve are well-investigated (A. Chakraborty, Baowaly, Arefin, & Bahar, 2012; Jacobson, Booch, & Rumbaugh, 1999; Nurmuliani, Zowghi, & Powell, 2004). In short, a clear, simplified, well-structured, and rationalized account of what is required can increase the possibility of a good system design (Mund et al., 2015), even if it involves accepting uncertain or conflicting opinions as definite requirements.

While there remains significant debate as to whether requirements should be clearly understood before system development begins (e.g., Kruchten, 2004) or as the development progresses (e.g., Beck, 2005; L. Cao & Ramesh, 2008), most agree that requirements will eventually be critical for any software engineering success (Verner, Cox, Bleistein, & Cerpa, 2005). Here, software engineering success refers to the net impact of a software system on its stakeholders over time (Ralph & Kelly, 2014). In contrast, research in inter-related fields of human-computer interaction (HCI), human-centred design and interdisciplinary research in design literature tends to assume that any design artefact has multiple stakeholders who often do not agree on either what problem an artefact should solve or how the artefact would solve the problem, or
both (Checkland & Scholes, 1999). Similarly, one of the main themes to emerge from behavioural decision-making research is that “people’s preferences are often constructed in the process of elicitation” (Lichtenstein & Slovic, 2006, p.1 ). Supporting this notion, empirical research views the design project as a system of inquiry, where a designer collaboratively constructs desiderata with stakeholders and produces real or imaginary solutions by building temporary design artefacts (Ralph, 2010b, 2015; Schon, 1984).

Analogously, one can consider RE to be creative process (Maiden & Gizikis, 2001; Maiden et al., 2010) in which multiple stakeholders and analysts collectively make sense of the problem situation and co-construct a shared mental model of a possible system (S. Chakraborty, Sarker, & Sarker, 2010). Expert designers usually reject the importance of the initial task structure (e.g., requirement specifications) and tend to proceed via an improvised and solution-focused exploration approach (Cross, 2001) that involves using ill-structured or ambiguous problem frames such as desiderata, ideas, or conjectures. Here, desiderata are properties of a real or an imaginary system that are wanted or needed by one or more stakeholders (Brooks Jr, 2010), and framing, a cognitive bias, is defined as “the tendency to give different responses to problems that have surface dissimilarities but are formally identical” (Stanovich, 2009, p. 88).

RE research has documented and summarized many different ways to present or communicate desiderata, such as personas (Haikara, 2007), a backlog of user stories (Schwaber, 2004), scenarios (Pohl, 2010), use-case narratives (Cockburn, 2000), and requirements specification statements (IEEE, 1998). These presentation techniques can be considered as various ways in which a problem is framed. When the desiderata are presented as compulsory or mandatory, they are being framed as “requirements”. In other words, a requirement can be considered as a property that a real or an imagined system must possess (Ralph & Wand, 2009).

However, the way in which a problem is presented or framed can also have significant effects on human cognition and the creative performance of designers (Ralph & Mohanani, 2015; Ward, Patterson, & Sifonis, 2004). Additionally, myriad empirical research shows that task structuring is negatively associated with design performance (Ralph & Mohanani, 2015). Specifically, “over-concentration (over-structuring) on problem definition does not necessarily lead to successful design outcomes” (Cross, 2004, p. 439). Some research has even suggested that framing the context of a software problem in terms of the “requirements” can seriously affect design creativity (Ralph, 2013).
In other words, formalizing and over-structuring desiderata as “requirements” may impede design performance, especially for non-critical software systems where creativity and innovation make major contributions to the project’s success. Surprisingly, SE research has not empirically investigated the effects of presenting or framing desiderata in different ways on creativity of design outcome. Misinterpreting all ostensible desiderata as compulsory or mandatory may interfere with designers’ ability to produce creative designs. Given explicit requirements, designers may fixate on initial solution ideas, experience, or available solutions and create less original designs (Ralph, 2016). Here, fixation, a cognitive bias, refers to the tendency to “disproportionately focus on one aspect of an event, object, or situation, especially due to self-imposed or imaginary obstacles” (Ralph, 2011b, p. 5).

This dissertation, therefore, investigates the effects of specification formality on design creativity to address this gap in SE research. Here, “specification formality” refers to the degree to which a problem statement is framed clearly, precisely, and in a structured way, whereas, “creativity” is conceptualized as a cognitive process that produces artefacts that are not only original but also practical (Runco & Jaeger, 2012).

1.1 Research gaps

Before investigating the effects of presenting desiderata as formal requirements on design creativity, it was necessary to explore the state of the art research concerning cognitive biases (including framing effects and fixation) in SE research and how creativity is understood and quantified in SE. Consequently, this research addresses the following research gaps:

**Research Gap 1: Lack of a comprehensive body of knowledge that summarizes the available literature in SE research involving cognitive biases, particularly concerning fixation and the framing effect.**

Investigations concerning human cognitive biases have only very recently gained popularity in SE research. There has been a constant increase in the number of publications investigating the effects and causes of cognitive biases across diverse SE activities. Despite the growing number of publications, there is still no comprehensive body of knowledge that attempts to systematically summarize the current state of SE research involving cognitive biases (before our published systematic mapping study (SMS) (Mohanani, Salman, Turhan, Rodríguez, & Ralph, 2018)). However,
inquiries concerning cognitive biases appear disconnected from one another in multiple ways—studies investigate the same biases in complete isolation without any reference to other relevant studies, the same biases are defined differently in different studies, and the inconsistent use of terminologies to explain cognitive biases create a great deal of unnecessary confusion. Therefore, it is difficult to follow and make sense of the discrete knowledge available across many studies and utilize this knowledge for future research investigations.

In the context of this dissertation, it was essential to explore the current state of SE research concerning two specific cognitive biases—fixation and the framing effect. Investigations related to them tend to follow a similar pattern as investigations concerning any other biases in SE, where biases are conceptualized differently and studies are disconnected from one another. Therefore, it was critical to identify how these biases manifest themselves in diverse SE activities and obtain a clear understanding of their causes and effects. Hence, systematically summarizing and aggregating the available knowledge in SE research will help to identify existing knowledge involving cognitive biases, recognize interesting patterns, and produce new knowledge by connecting, integrating, and interpreting together previously unconnected research. Moreover, similar reviews have been conducted in the field of information systems (IS) that provide a general overview of cognitive biases (e.g., Arnott, 2006; Ralph, 2011a). A systematic literature review (SLR) conducted by Fleischmann, Amirpur, Benlian, and Hess (2014) further identified areas in IS that have been actively investigated concerning cognitive biases and areas where additional investigations are warranted. However, no analogous review exists in SE.

**Research Gap 2: Lack of shared agreement between SE research and practice concerning the way in which creativity is understood and assessed in SE.**

In SE, creativity has mostly been investigated in the context of RE (Nguyen & Shanks, 2009), agile development (Crawford, de la Barra, Soto, & Monfroy, 2012), software design (Horowitz, 1999), and open-source software development (Dexter & Kozbelt, 2013). However, very few SE studies attempt to understand creativity in SE contexts (e.g., Crawford et al., 2012; Maiden et al., 2010), (Graziotin, Wang, & Abrahamsson, 2014) or try to investigate the importance of creative performance from the practitioners’ perspective (e.g., Fischer, Boogaard, & Huysman, 1993; Sutling, Mansor, Widyarto, Letchmunan, & Arshad, 2014). Instead, most of the studies about
creativity in SE tend to propose tools and techniques to enhance creativity (e.g., Bartelt, Vogel, & Warnecke, 2013; Díaz, Aedo, Rosson, & Carroll, 2010).

Research concerning creativity can be broadly classified into the “Six P’s” (Rhodes, 1961; Sternberg, 1999), as follows:

– creativity as a cognitive process
– the creative product
– the person doing creative work
– the place (context) of creative work
– ways to enhance creative thinking (persuasion) and
– ways to improve creative potential.

The main objective of this dissertation involves assessing the creativity levels of the solution designs (i.e., evaluating the creativity of a product). However, assessing creativity remains a challenge, as creativity is subjective, multidimensional, prone to confusion in conceptualization, and difficult to understand and quantify. Nonetheless, only a shared understanding of how creativity is understood in SE can inform creativity assessment techniques (T. M. Amabile, 1983).

While researchers and practitioners in psychology and philosophy have built a common understanding around the various dimensions of creativity in their domains (e.g., Sternberg, 1999), SE practitioners and researchers have not. Moreover, the role of creativity in SE can be better understood only by taking into account the understanding, expectations, and reservations of both SE researchers and practitioners. Finally, by comparing and contrasting researchers’ and practitioners’ perspectives on creativity, one can better understand and explore the obstacles to knowledge transfer and the existence of any gap between SE research and practice.

Research Gap 3: The question of whether or not more structured desiderata framing (i.e., specification formality) impedes design creativity.

Interdisciplinary research, both empirical and theoretical, concerning problem-structuring, goal-understanding, creativity, and sensemaking in psychology, sociology, and design science has argued that task structuring—which involves assigning socially constructed constraints, opinions, parameters, states and other specifics to the context (Dorst & Cross, 2001)—has a negative relationship with design performance (e.g., Dorst, 2006; Ralph, 2011a). In other words, well-structured and over-rationalized tasks often lead to poorer design performance and less software design success. Moreover,
myriad empirical studies show that solving ill-structured or ambiguous problems actually improves design performance, as follows: 1) unclear or “open” goals lead to more learning effects (Vollmeyer & Burns, 2002; Wirth, Künsting, & Leutner, 2009) and increase unconscious assimilation of context-relevant information (Moss, Kotovsky, & Cagan, 2007); 2) more ambiguous task framing leads to more original solutions than more specific task framing (Butler, Scherer, & Reiter-Palmon, 2003; Ward, Patterson, & Sifonis, 2004); and 3) accepting conflicting objectives and uncertainty can lead to more effective solutions (Butler et al., 2003). All this evidence combines to suggest a negative relationship between greater task structure and the quality of design performance.

This research gap, therefore, aims to study whether more structured and formal framing of desiderata as requirements statements can reduce design creativity.

1.2 Research objective and research questions

The main objective of this dissertation is to empirically test and explore the effects of framing desiderata as formal requirements statements on creative design performance. Consequently, the main research question that guides this dissertation is as follows:

Research Question: What is the effect of specification formality on design creativity?

Based on the research gaps discussed in Section 1.1, the main research question is further divided into the following sub-research questions:

RQ1: What is the current state of research concerning cognitive biases and creativity in SE?
- RQ1.1: What is the current state of the art in SE research involving cognitive biases?
- RQ1.2: How is creativity conceptualized, assessed, and improved in SE?

RQ2: How does specification formality affect design creativity?
- RQ2.1: Does the framing of desiderata affect design creativity?
- RQ2.2: What cognitive behaviours can explain the relationship between specification formality and design creativity?

While RQ1 addresses research gaps 1 and 2, RQ2 addresses research gap 3. RQ1 is conceptualized to help identify the current state of the art in SE research concerning
cognitive biases and to understand how researchers and practitioners perceive creativity. RQ2 investigates the impact of presenting desiderata as formal requirements statements on design creativity and explores the cognitive mechanisms underlying these causal effects. A summary of the research gaps, research questions, and publications addressed in the thesis are shown in Figure 1.

Fig. 1. Mapping of research gaps to research questions and publications.

1.3 Scope of the research

The main focus of this research is to demonstrate the harmful effects of framing ostensible desiderata as *requirements* statements on design creativity. That is, to investigate the possibility that misinterpreting all requirements as mandatory might interfere with the designer’s ability to produce creative artefacts. The evidence supporting the relationship between desiderata framing and creativity is based on controlled experiments. This evidence is triangulated by performing a qualitative study (i.e., dialog-based protocol study) (Xu & Rajlich, 2005) not only to further explore the relationship but also to explain the theoretical principles underlying the causality between the variables. Hence,
the research approach in this dissertation can be described as a mixed method research design (Creswell & Clark, 2017), as it involves collecting, analysing, and interpreting both quantitative and qualitative data in a series of studies. While controlled experiments are used to collect quantitative data, the protocol analysis method is used to collect and analyse qualitative data to fulfil the main objective of this dissertation. Due to the limited temporal scope of doctoral training, the empirical data were collected by conducting short-term laboratory experiments and a protocol study instead of a long term longitudinal study. The findings from both the controlled experiments and protocol study were interpreted together to answer the main research question of the dissertation.

However, to understand how cognitive biases such as fixation and the framing effect manifest themselves in various SE activities and to identify the relevant SE literature to guide further investigations, an SMS was conducted to aggregate and summarize the available knowledge concerning cognitive biases in SE. Similarly, to investigate how creativity is conceptualized and evaluated in SE, an SMS was combined with interviews involving practitioners to explore any shared agreement (or lack thereof) in the way creativity is perceived by SE researchers and practitioners.

The operationalization of the results—developing debiasing techniques to mitigate fixation on desiderata due to the framing effect—is beyond the scope of this dissertation. Any debiasing technique or approach needs to be empirically validated before being used in an actual context; untested debiasing techniques may have unintended consequences.

The particular dimensions of creativity under investigation, originality and practicality of designs, were chosen based on the rationale presented in Chapter 2. Investigating the effects of desiderata framing on other dimensions and factors that influence creativity is beyond the scope of this dissertation. These include, but are not limited to, motivation (T. Amabile, 1996), domain knowledge (Baer & Kaufman, 2005; Kilgour, 2006) and affective states (Frijda, 1987; Isen, Daubman, & Nowicki, 1987). Moreover, certain other dimensions of product creativity (e.g., simplicity, marketability, ease of use), appear to conflate creativity with general quality attributes of software and are beyond the scope of the investigations reported in this dissertation.

Finally, the scope of this research is limited to the point of view of software designers and developers, although the results of this dissertation and the topic of investigation—the effect of cognitive biases and the importance of creativity in SE practices—could constitute a baseline to be taken into consideration by other stakeholders, such as project managers, testers, and SE researchers.
1.4 Dissertation structure

The dissertation is organized into six chapters. Following the introduction, Chapter 2 reviews the existing literature on fixation, framing effect, task structuring, design thinking, and creativity and provides essential definitions of the key terms for positioning this empirical investigation within the broader research area of SE. Chapter 3 presents the research framework, including detailed descriptions of the various research methods used to address the research gaps. Chapter 4 presents the original publications included in this dissertation. Each contribution is based on the research questions that guide this work. Chapter 5 summarizes the main results by answering the individual research questions and discusses various threats to validity. Finally, Chapter 6 concludes with a summary of the contributions and the implications of the results on SE practice and education before identifying future research opportunities.
2 Background and related work

The work presented in this dissertation is positioned around the areas of software design, cognitive biases (i.e., fixation and framing effects), and design creativity.

Section 2.1 provides a brief overview of cognitive biases, focusing on fixation and framing effect. This is followed by a brief discussion about task structuring and design thinking in sections 2.2 and 2.3 respectively. Finally, section 2.4 discusses the concept of creativity within the scope of software engineering.

2.1 Cognitive biases

One of the major research areas in SE research concerns human behaviour and cognition (de Souza, Sharp, Singer, Cheng, & Venolia, 2009; Sage & Rouse, 1999; Zsambok & Klein, 2014). Many studies concerning human decision-making in SE adopt theories and concepts from psychology to address issues in trivial SE problems like decision support systems (Arnott, 2006; Arnott & Pervan, 2008). One of these concepts is cognitive biases, which are systematic deviations from optimal reasoning (Ralph, 2011b). More generally, cognitive biases are frequent errors in thinking or patterns of bad judgment that are observed in different people and different contexts (Arnott, 2006). Cognitive biases help explain many problems occurring in myriad SE activities, including software design (Mair & Shepperd, 2011), requirements engineering (Chotisarn & Prompoon, 2013), software testing (Calikli & Bener, 2010), and software project management (Jorgensen & Grimstad, 2012).

SE research on cognitive biases helps to identify common errors, their causes, and ways to mitigate these errors. Additionally, the concept of cognitive biases is also used to create better SE practices (Browne & Ramesh, 2002; Calikli, Bener, Caglayan, & Misirli, 2012), methods (Mohan & Jain, 2008; Snow, Keil, & Wallace, 2007), and artefacts (Parsons & Saunders, 2004). Some examples of how cognitive biases affect SE activities are as follows:

- **confirmation bias** is used to explain the common antipattern where unit tests attempt to confirm that the code is successful rather than to reveal code failures (Calikli & Bener, 2010)
- **the framing effect** affects design creativity (Mohanani, Ralph, & Shreeve, 2014)
optimism bias affect all kinds of software projects, leading to delays and budget overruns (Flyvbjerg, 2006; Jørgensen, 2010).

All SE activities require practitioners to process large amount of information spread across the problem and solution domains (Clayton, Rugaber, & Wills, 1998). Consequently, research on cognitive biases is critical to making sense of this knowledge with the aim of improving SE practices. However, identifying the causes of cognitive biases is challenging. Most research is concentrated on investigating heuristics, the simple rules of thumb for making decisions and judgments. In this regard, some of the most investigated heuristics include anchoring and adjustment—estimating a quantity by adjusting from an initial starting value (VandenBos, 2007)—the availability heuristic—estimating the importance of something based on how easy it is to remember (Poses & Anthony, 1991). Cognitive biases can even be triggered by other innate phenomena like emotions (Pfister & Böhm, 2008), social interactions and influences (Yechiam, Druyan, & Ert, 2008), and noisy information processing (Hilbert, 2012). Despite all this research, the cognitive mechanisms underlying many cognitive biases and their manifestation remain under-investigated.

Along with the causes and effects of cognitive biases, SE research also investigates ways to prevent biases or mitigate their harmful effects, a process known as debiasing (Larrick, 2008; Stacy & MacMillan, 1995). Fischhoff categorizes debiasing interventions into five levels: “To eliminate an unwanted behavior, one might use an escalation design, with steps reflecting increasing pessimism about the ease of perfecting human performance: (A) warning about the possibility of bias without specifying its nature; (B) describing the direction (and perhaps extent) of the bias that is typically observed; (C) providing a dose of feedback, personalizing the implications of the warning; (D) offering an extended program of training with feedback, coaching, and whatever else it takes to afford the respondent cognitive mastery of the task” (1981, p. 426). However, the issue with these levels is that approaches A, B, and C are seldom effective (Pronin, 2007; Pronin, Lin, & Ross, 2002), and option D can become very expensive. Fischhoff, therefore, proposed another option—to debias the task instead of the individual. For example, Planning Poker, a technique used in many agile methods, redesigns effort estimation to prevent developers from anchoring on the initial estimates (Haugen, 2006).

The dissertation is primarily concerned with fixation and framing effect.
2.1.1 Framing effect

Sometimes, the manner in which decision options are presented can bias the decision-making process (Hastie & Dawes, 2010). One such bias is the framing effect, defined as “the tendency to give different responses to problems that have surface dissimilarities but that are really formally identical” (Stanovich, 2009). These frames can be considered cognitive shortcuts that individuals use to assimilate and understand complex information and to simplify this complex information into more easily understandable components (Liu, Wang, & Zhao, 2010).

For instance, participants in an experiment were asked to select one of the two treatments for a hypothetical disease—treatment A, which would save 200 of 600 people, or treatment B, which had a 1-in-3 chance of saving everyone and a 2-in-3 chance of saving none. Participants were thus asked to choose between 400 people definitely dying or a 33 percent chance that nobody would die. Most participants chose the latter option. However, when participants were asked to choose between definitely saving 200 people or a 1-in-3 chance of saving everyone, most participants chose the former (Tversky & Kahneman, 1985).

Here, the difference in response is attributed to the manner in which the question are framed rather than the actual facts. The framing effect is extremely robust (Bohm & Lind, 1992; Cheng & Wu, 2010) and affects different individuals in diverse circumstances (Levin, Schneider, & Gaeth, 1998).

2.1.2 Fixation

Fixation occurs when people “disproportionately focus on one aspect of an event, object or situation, especially self-imposed or imaginary obstacles” (Ralph, 2011b, p. 5). Fixation is one of the many biases (including anchoring and adjustment heuristic and representativeness) related to selective perception—the tendency for different people to perceive the same scenarios differently (Plous, 1993)—and is typically used to explain and investigate the mental processes of designers during early stages of the design process. Jansson and Smith (1991) have conducted several experiments to demonstrate design fixation—the tendency for designers to generate solutions. These experiments were repeated by providing designers with very similar examples (Youmans & Arciszewski, 2014) and existing designs or artefacts (Finke, 1996). These experiments hypothesized that designers would alternate between the “configuration space”—a
hypothetical (mental) space that contains physical representation of the artefact in the form of its design, diagrams, or sketches—and the “concept space”—a hypothetical (mental) space where abstract ideas and relationships or patterns between the ideas are considered. Any barriers faced by designers that affect their ability to think between these two spaces would impede the process of conceptual design. Simply stated, fixation prevents designers from finding novel solutions or original associations or envisioning alternative ideas (Zahner, Nickerson, Tversky, Corter, & Ma, 2010).

Although providing examples causes design fixation, the effects of fixation are moderated in several ways, as follows:

- the magnitude of fixation varies by the domain as when, for example, mechanical engineers fixate more than industrial engineers (A. Purcell, Williams, Gero, & Colbron, 1993; A. T. Purcell & Gero, 1996)
- common, typical examples induce more fixation than rare examples (Perttula & Sipilä, 2007)
- fixation can be reduced by explicit “de-fixating” instructions, which are sets of guidelines to avoid common features of existing examples (Chrysikou & Weisberg, 2005)
- providing high-quality examples can lead to more creative designs than providing low-quality or no examples at all (Lujun, 2011)
- product dissection activities (Toh, Miller, & Kremer, 2014) and physical prototyping (Youmans, 2011) can mitigate fixation
- specifically in SE, ambiguity in requirements specifications may reduce premature fixation (Nuseibeh & Easterbrook, 2000), and how the task is presented may also inform fixation (Kim & Ryu, 2014; Zahner et al., 2010).

This last point indicates the relationship between fixation and the framing effect, and many studies investigating fixation have leveraged framing effect (Kim & Ryu, 2014; Perkins, 2009). Furthermore, earlier research (e.g., Atilola, Tomko, & Linsey, 2016; Chrysikou & Weisberg, 2005) has investigated fixation by providing designers with varied instructions. In these studies, the independent variable (IV) was conceptualized as task framing, and the resulting fixation was conceptualized to have occurred due to the framing effect. This dissertation extends the idea that designers fixate on given examples or early ideas, to hypothesizing that task structuring causes fixation.
2.2 Task structuring

Problems are often conceptualized on a spectrum of well to ill structured. “Well-structured problems are constrained problems with correct or convergent solutions that require the application of a limited number of rules and principles within well-defined parameters; whereas, ill-structured problems possess multiple solutions and fewer parameters that are less manipulable and contain uncertainty about the concepts, rules, and principles that are necessary for the solution, the way they are organized and which solution is best” (Jonassen, 1997, p. 65). Although most design problems are ill structured in nature (Newell & Simon, 1972; Simon, 1996), “problem-solving theory that is based upon the solution of well-structured problems should serve as the basis for all problem solving” (Dorst, 2006, p. 7). However, interdisciplinary empirical research directly contradicts this common notion (Dorst, 2006; Ralph, 2011a).

Moreover, empirical studies across many domains have demonstrated that a greater task structure—providing more structure, clarity, and certainty to a problem—impedes design performance for the following reasons:

- more abstract or non-specific task framing lead to more creative solutions (Ward, Patterson, & Sifonis, 2004), and presenting conflicting opinions or objectives results in more effective solutions (Butler, Scherer, & Reiter-Palmon, 2003)
- non-specific or open goals reduce cognitive load, which leads to a more powerful learning effect (Wirth, Künsting, & Leutner, 2009)
- designers often tend to fixate on their experience (Jansson & Smith, 1991) on existing solutions (A. T. Purcell & Gero, 1996) or on initial set of (early) ideas (Guindon, 1990)
- designers tend to process what little information they acquire and quickly translate this information into the problem schema (i.e., configuration space), which improves their understanding of the problem (Kruger & Cross, 2006).

To summarize, all these points suggest a negative relationship between task structuring and creative design performance outcome. Specifically, “over-concentration on problem definition does not lead to successful design outcomes” (Cross, 2004, p. 439).

On the other hand, traditional RE practices advocate the idea of providing more structure to a problem, which is known as “problem structuring” (Ralph & Mohanani, 2015). Problem structuring involves human agents assigning socially constructed constraints, personal opinions, choices, states, and other specifics to a problem context.
(Dorst, 2006). This notion has motivated a wide range of practices and techniques to present and communicate desiderata more clearly as user stories (Schwaber, 2004), scenarios (do Prado Leite et al., 1997), use-case narratives Cockburn (2000), and requirements specifications (IEEE, 1998). Prioritizing requirements adds further structure and clarity. In short, it appears that a clearer, more rationalized, and more structured account of what is required or demanded by the user would improve the likelihood of project success. The present study investigates the possibility that more structured tasks could lead to poorer design performance and less success.

In the context of this dissertation, presenting desiderata as requirements is conceptualized as providing structure, clarity, and completeness to a problem situation. On the other hand, presenting desiderata as ideas is considered more uncertain, unclear, and ambiguous than requirements framing. Prioritizing requirements adds further clarity, certainty, and structure to the desiderata.

2.3 Design thinking

This section summarizes the concept of software design and design thinking according to the empirical paradigm’s model of design (Ralph, 2010a). Here, software design refers to the process where an agent or a group of agents create a software artefact, intended to accomplish goals in a specific context by using a set of primitive components and implementing a set of requirements, subject to a set of constraints (Ralph & Wand, 2009).

Design problems are often regarded as ill-structured problems (Newell & Simon, 1972), where the process of “doing design” is explained by the constructivist view of human perception and thought process known as reflection-in-action (RiA). According to the RiA view, a “design” is conceptualized as a reflective conversation during which the designer alternates between framing the problem and executing a set of actions intended to address the problem and the solution together (i.e., making moves) and simultaneously evaluating those moves (Schon, 1984). This evaluation can either result in a satisfactory solution, a re-framing of the problem conceptualization, or even a complete change in one’s perceptions of the given design task (Dorst, 1997).

In this way, a design is perceived as a kind of problem-solving process, in which a designer, based on his/her assessment of the situation, creates early solution conjectures, continuously refines the problem and the solution situation, and uses the conjectures as evanescent ideas to explore and define the problem and solution together. These
strategies that help a designer to understand, synthesize, and creatively convert the knowledge to generate original design concepts are collectively known as design thinking (Cross, 2011).

Furthermore, design thinking can be better explained by differentiating between the “problem space”—the metaphoric space comprised of the mental representations of the designers’ interpretation of the problem situation—and the “solution space”, the space that contains the mental representation of the designers’ interpretation of the solution to the problem (Purao, Bush, & Rossi, 2001). This process could be understood more simply as designers trying to understand the problem by generating initial ideas about the (possible) solution. By doing so, designers enhance their understanding of the problem, which in turn triggers new and creative solution ideas. This process, in which a designer acts as a creative agent who oscillates between an ill structured problem definition and a tentative solution idea to simultaneously and gradually improve both, is known as co-evolution (Cross, 2011). More generally, co-evolution is the process where design agents mutually explore and enhance their perception of the problem space and the design space (Dorst & Cross, 2001). Co-evolution refers to a process where two or more inter-related objects mutually change and develop over time in such a way that any change in one is used to inform the changes in the other (Ralph, 2016).

2.4 Creativity

Although no single definition of creativity is universally accepted, it has been referred as “a process of becoming sensitive to problems, deficiencies, gaps in knowledge, missing elements, disharmonies, and so on; identifying the difficulty; searching for solutions, making guesses, or formulating hypotheses about the deficiencies; testing and retesting these hypotheses and possibly modifying and retesting them; and finally communicating the results” (Torrance, 1966, p. 6). However, Amabile’s definition of creativity as the “production of novel and useful ideas in any domain” (T. Amabile, 1996) is the most widely accepted conceptualization of creativity (Sosa & Danilovic, 2009; Sternberg, 1999). In other words, a creative artefact is not only novel (Mayer, 1999; Plucker & Makel, 2010) but should also be appropriate for actual use in a specific context (Christiaans, 2002; Mumford, 2003; Runco & Jaeger, 2012). Moreover, this dual criterion of creativity is well accepted in engineering design (Sosa & Danilovic, 2009) and software engineering (Maiden, Manning, Robertson, & Greenwood, 2004).
The dual view of creativity suggests the assessment of creative artefacts by evaluating the originality and practicality of an artefact (Christensen & Ball, 2016; Shah, Smith, & Vargas-Hernandez, 2003). When assessing creativity, not all criteria are equal. The “usefulness” of the artefact is only considered when the solution is novel (Diedrich, Benedek, Jauk, & Neubauer, 2015). Additionally, objective creativity metrics remain under-investigated and highly contested (Chi, 2006; Dillon, 1982). Therefore, the most common creativity assessment technique is subjective assessment by expert judges (Adams, 2005; Baer & Kaufman, 2005; Runco & Charles, 1993). However, judges often do not reach a consensus, which can undermine an evaluation’s validity (Plucker, 2005).

Creativity is a multidimensional construct (T. M. Amabile, 1988) that is often linked with divergent thinking (Guilford, 1965)—exploring multiple alternative solutions to a problem. In SE research specifically, creativity workshops are frequently organized to provide clarity, enhance the quality of requirements specifications (Crawford, de la Barra, Soto, & Monfroy, 2012; Maiden et al., 2010), and produce novel requirements (Wen, Zhang, Liu, & Yang, 2010). Interactive collaboration techniques are used by stakeholders to produce highly creative requirements (Horowitz, 1999), so software requirements can be considered an encapsulation of creative thinking about the system (Maiden, Manning, Robertson, & Greenwood, 2004); since RE is a highly creative process (Maiden & Gizikis, 2001; Maiden et al., 2010), it can be improved by incorporating creativity-enhancing techniques (Maiden, Manning, et al., 2004).
3 Research design

The work presented in this dissertation follows a combination of explanatory and exploratory research approaches that involve controlled experiments, semi-structured interviews, and a protocol study. While controlled experiments were used to show the relationship between the variables, semi-structured interviews and protocol analysis were used to seek out and gain insights about the phenomenon under investigation. This research also used secondary reviews to aggregate and synthesize the existing literature related to a specific topic with the aim of integrating the evidence in the light of predefined research questions. Section 3.1 provides a description of the main research phases of this dissertation, with a detailed description of each research method used in the context of the dissertation explained in section 3.2. The motivation for using various research methods is as follows:

1. The main objective of this dissertation was to investigate the effect of presenting desiderata as formal requirements statements on creative design performance. In other words, this research aims to identify and explain the effect of framing desiderata as formal requirements on design creativity. First, an SMS was conducted to understand the state of the art research involving cognitive biases in SE at a high level of granularity (Petersen, Feldt, Mujtaba, & Mattsson, 2008). In the specific context of this dissertation, one of the sub-goals of this secondary review was to understand how fixation and the framing effects are manifested in SE activities and if any existing research investigates the relationship among framing effect, fixation, and design creativity. Additionally, an SMS would help reveal other cognitive biases that affect creativity in the context of specification formality, which could lead to additional insights explaining this phenomenon. The overall research design for SMS is described in section 3.2.1.

2. It was also necessary to develop a clear understanding of the role of creativity in SE, with a particular focus on the way creativity is conceptualized and assessed in SE. As only a shared understanding of the concept of creativity between SE research and practice could guide the creativity assessment, an SMS was combined with an interview study that compared and contrasted the way creativity is perceived by SE researchers and practitioners. A more detailed explanation concerning the execution of interviews appears in section 3.2.2.
3. An explanatory research approach was then used to establish the relationship between specification formality and creativity by using controlled experiments (Wohlin et al., 2012). An explanatory study implies that the research aims to explain rather than describe a phenomena (Yin, 2017). The controlled experiments tested prior hypotheses by quantitatively measuring the variables of interest. The data were analysed using statistical techniques. Since there is no a-theory explaining the effect of desiderata framing on creativity, the hypotheses were constructed based on the theoretical conceptualization of fixation and creativity (see Chapter 2).

4. Contrary to explanatory research, exploratory research aims to describe or gain a deeper understanding of the phenomena under investigation, usually via qualitative data collection methods. While quantitative analysis provides systematic evidence of the relationship between the dependent variable (DV) and IV, qualitative analysis tends to comprehend better the personal perspectives and experiences of the individuals involved. To this effect, a dialog-based protocol study (Xu & Rajlich, 2005) was conducted to explore the cognitive mechanisms that could explain the underlying causal effects between the variables. The qualitative data were analysed thematically (Cruzes & Dyba, 2011) by adopting an appropriate coding scheme, as proposed by Saldana (2015).

3.1 Research phases

The overall research design is illustrated in Figure 2. The research presented in this dissertation was carried out three phases:

3.1.1 Phase 1: Identifying the relevant literature and variables; Papers I and II

The first step in any research endeavour is to identify the literature and a practical and relevant problem. In this phase, the existing literature concerning cognitive biases investigated in SE was systematically reviewed, with a particular focus on how fixation and framing effect manifest themselves in SE activities. Secondly, a systematic review of existing SE literature concerning creativity was combined with interviews involving SE practitioners to investigate and explore the role of creativity and the importance of creative performance in SE. This phase thus identified the IV and DV used in subsequent empirical investigations, the hypothesized relationship between the variables
(as conceptualized by existing theoretical evidence), and salient factors that constitute the research gaps. Additionally, the hypotheses that guided the experimental work were formulated during this phase. It should be noted that the literature on the topic was regularly reviewed during the subsequent phases to keep it up to date. The findings from this phase guided and informed the subsequent empirical research conducted in phases 2 and 3.

3.1.2 Phase 2: Experimentation; Papers III and IV

Using controlled experiments, Phase 2 investigated the hypothesized effect of specification formality on design creativity. Specifically, two between-subjects randomized controlled trials were conducted to compare the differences in design creativity levels due to desiderata framing. During this phase, the author conceptualized and developed the task and other relevant materials used to conduct the experiments. Participation in both the experiments was solicited from a population of post-graduate students (section 3.2.3 provides a more detailed description of this process).

3.1.3 Phase 3: Protocol study; Paper V

While the experiments conducted in phase 2 were used to establish causality between the variables (i.e., specification formality and design creativity), the main objective of phase 3 was to explore the underlying cognitive processes that could explain the relationship between the variables. This process involved conceptualizing, planning, designing, and executing a dialog-based protocol study (Xu & Rajlich, 2005) that explored the cognitive processes underlying the causal effects. This study was executed by soliciting participation from software practitioners and post-graduate students.

In this dissertation, RQ1 is answered by using the findings from Paper I and Paper II, while RQ2 is answered using the empirical evidence found in Papers III, IV, and V.

3.2 Research process

This section provides a detailed description of the research methods (including data collection and analysis procedures) used to fulfil the objective of this dissertation.
A replication package containing all the documents used to conduct the interviews, experiments, and protocol study reported in Papers II, III, IV, and V is available in Appendix 1.

### 3.2.1 Systematic mapping study (SMS)

An SMS is a form of secondary literature review that uses similar methodological steps as a systematic literature review, with the differences lying in the synthesis of research outcomes. Specifically, an SMS classifies the primary studies based on various bibliographic information and other topic-specific themes that are relevant to the chosen topic of interest. In general, an SMS covers a broader topic, while a systematic literature review has a rather focused research topic area (Kitchenham, Brereton, Li, Budgen, & Burn, 2011). To address RQ1.1, an SMS was adopted because the existing research on cognitive biases in SE was fragmented and did not follow common terminology or use established theoretical concepts. Moreover, aggregating and categorizing the available research on creativity in the SE research literature (RQ1.2) appeared quite broad, so an SMS was preferred over a detailed review. An SMS approach works well in such circumstances because it enables categorizing and aggregating the available knowledge dispersed across many disconnected studies at a high level of granularity.

The SMS approach in Papers I and II followed the established guidelines in (Kitchenham,
Defining research questions

For Paper I, the research questions of the SMS were conceptualized by the complete author team based on the research purpose—to review, summarize, and synthesize the current state of software engineering research involving cognitive biases. Based on the research objective, the research questions guiding this SMS were defined as follows:

– What cognitive biases are investigated?
– What antecedents of cognitive biases are investigated?
– What effects of cognitive biases are investigated?
– What debiasing approaches are investigated?
– What research methods are used?
– When were the articles published?
– Where were the articles published?
– Who is conducting the research?
– Which software engineering book of knowledge (SWEBOK) knowledge areas are targeted by research on cognitive biases?

The main objective of Paper II was to understand the role of creativity in SE, with a particular focus on how creativity is conceptualized, assessed, and improved in SE. The research questions were defined as follows:

– How is creativity conceptualized in SE?
– What factors influence creativity in SE?
– How is creativity measured in SE?

Primary study selection process

While the mapping protocol for SMS conducted in Papers I and II consisted of five main steps—defining the research questions, searching online databases for primary studies, screening the retrieved studies for inclusion and exclusion criteria, evaluating study quality, and extracting the relevant data—the mapping process in Paper I required additional steps, mainly to ensure sampling adequacy. These additional steps were applying backward snowballing, reviewing the publication list of the most active authors,
conducting an unfiltered search on Google Scholar\(^1\), and searching for specific cognitive biases. While a detailed description of the primary study selection process for both studies can be found in the original publications of Papers I and II, Table 1 summarizes the main results from each step of the screening procedure conducted in both papers.

**Data extraction and mapping**

The data elements in both studies were extracted using the qualitative data analysis software NVIVO\(^2\), which is good for organizing, analysing, and visualizing unstructured qualitative data and for classifying primary studies based on the relevant bibliographic information. The primary studies were coded deductively (Saldaña, 2015) using an *a priori* list of codes that were developed by the author based on the set of pre-defined research questions. The codes were then organized appropriately to answer the research questions. The coding schema and subsequent classifications for both studies were reviewed by the entire author team. The results of the respective mapping studies and their contributions to this dissertation are summarized in Sections 4.2 and 4.3.

**3.2.2 Cross-sectional interviews**

RQ1.2 was partially answered in Paper II by conducting constructivist interviews to explore software practitioners' perspectives on creativity. A semi-structured interview guide with open-ended questions was developed collaboratively by the entire author team based on established guidelines (DiCicco-Bloom & Crabtree, 2006; Whiting, 2008). The interviews addressed the following research questions from different directions—*how is creativity conceptualized in SE?*, *what factors influence creativity in SE?*, and *how is creativity measured in SE?*.

**Data collection**

Before the actual data collection process, the interview guide was trialled in three pilot interviews. After a few minor changes to improve the clarity of the questions, the data were collected by interviewing a convenience sample of 17 software practitioners, comprising 16 males and 1 female. Out of the 17 interviews (IC01–IC17), 5 interviews

---

\(^1\)scholar.google.com  
\(^2\)www.qsrinternational.com
Table 1. Summary of systematic mapping study reported in Papers I and II.

<table>
<thead>
<tr>
<th>Study selection steps</th>
<th>Paper I</th>
<th>Paper II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Research Question</td>
<td>What is the current state of the art in SE research involving cognitive biases?</td>
<td>How is creativity conceptualized, assessed, and improved in SE?</td>
</tr>
<tr>
<td>Main Data Elements</td>
<td>Cognitive biases, effects, antecedents, debiasing techniques, SE knowledge areas</td>
<td>Conceptualization of creativity, factors influencing creativity, creativity assessment techniques</td>
</tr>
<tr>
<td>Search String</td>
<td>“software” AND “cognitive bias”</td>
<td>“creativity” AND “software engineering”</td>
</tr>
<tr>
<td>Online Databases Used</td>
<td>IEEE Xplore, Scopus, Web of Science, ACM Digital Library and Science Direct</td>
<td>Scopus, IEEE Xplore, ACM Digital Library, Science Direct</td>
</tr>
<tr>
<td>Search Conducted Until</td>
<td>December 2016</td>
<td>November 2015</td>
</tr>
<tr>
<td>Primary Screening</td>
<td>40</td>
<td>84</td>
</tr>
<tr>
<td>Backward Snowballing</td>
<td>16</td>
<td>NA</td>
</tr>
<tr>
<td>Authors’ Analysis</td>
<td>9</td>
<td>NA</td>
</tr>
<tr>
<td>Unfiltered Search</td>
<td>0</td>
<td>NA</td>
</tr>
<tr>
<td>Search Specific Bias Name Articles Excluded (Quality Assessment)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Total # of Primary Studies</td>
<td>65</td>
<td>84</td>
</tr>
</tbody>
</table>

were conducted face-to-face and 12 via audio/video conference. Each interview lasted an average of 50 to 60 minutes. Participants had a mean experience of around six years in varying roles and industries (see Table 2). All interviews were audio-recorded and later transcribed for analysis, preserving the original grammar, verbal static and so on. The actual data collection process was carried out over three weeks from October 2015 to November 2015.

Data analysis

All interview transcripts were stored in NVIVO software for qualitative analysis. All the transcripts were coded line by line using an integrated coding approach (Couger, Higgins, & McIntyre, 1993). An integrated coding technique is a combination of inductive and deductive coding in which an initial set of codes based on the research
questions helps to develop new codes inductively. In software engineering research, thematic synthesis is a method used in systematic reviews to help identify recurring themes or patterns within primary studies that are then analysed to draw conclusions. In the present study, codes related to each pre-defined category were combined to form themes, with each theme was seen as a high-level conceptualization of multiple codes grouped together (Cruzes & Dyba, 2011). All codes and themes were iteratively checked and reviewed by the entire author team until complete agreement was achieved. The results from the qualitative analyses were compared and contrasted against the dimensions of creativity that emerged from the mapping study, as explained above.

### Table 2. Characteristics of the interviewees, ©2017 IEEE.

<table>
<thead>
<tr>
<th>ID</th>
<th>Position</th>
<th>Industry</th>
<th>Experience (In Years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IC01</td>
<td>Developer / Tester</td>
<td>Telecommunication</td>
<td>9</td>
</tr>
<tr>
<td>IC02</td>
<td>Developer</td>
<td>Software development</td>
<td>3.5</td>
</tr>
<tr>
<td>IC03</td>
<td>Developer</td>
<td>Software development</td>
<td>2</td>
</tr>
<tr>
<td>IC04</td>
<td>Developer</td>
<td>Telecommunication</td>
<td>10</td>
</tr>
<tr>
<td>IC05</td>
<td>System Architect</td>
<td>Marketing</td>
<td>9</td>
</tr>
<tr>
<td>IC06</td>
<td>Designer</td>
<td>Digital broadcast service</td>
<td>5.5</td>
</tr>
<tr>
<td>IC07</td>
<td>Researcher / Developer</td>
<td>Finance</td>
<td>5</td>
</tr>
<tr>
<td>IC08</td>
<td>Developer</td>
<td>Telecommunication</td>
<td>4.5</td>
</tr>
<tr>
<td>IC09</td>
<td>Developer / Tester</td>
<td>Software development</td>
<td>5</td>
</tr>
<tr>
<td>IC10</td>
<td>Developer</td>
<td>Marketing</td>
<td>10</td>
</tr>
<tr>
<td>IC11</td>
<td>Developer</td>
<td>Software development</td>
<td>3.5</td>
</tr>
<tr>
<td>IC12</td>
<td>Developer / Tester</td>
<td>Oil and gas</td>
<td>7</td>
</tr>
<tr>
<td>IC13</td>
<td>Developer / Tester</td>
<td>Oil and gas</td>
<td>6</td>
</tr>
<tr>
<td>IC14</td>
<td>Tester</td>
<td>Engineering and consulting</td>
<td>3.5</td>
</tr>
<tr>
<td>IC15</td>
<td>Tester</td>
<td>Software development</td>
<td>3</td>
</tr>
<tr>
<td>IC16</td>
<td>Developer / Tester</td>
<td>Software development</td>
<td>6.5</td>
</tr>
<tr>
<td>IC17</td>
<td>Tester</td>
<td>Telecommunication</td>
<td>12</td>
</tr>
</tbody>
</table>

### 3.2.3 Controlled experiments

In the context of this dissertation, two between-subjects randomized controlled trials were conducted in Papers III and IV, which investigated the relationship between desiderata framing and design creativity. More generally, the two experiments tested the hypotheses that more structured and formal specifications produce designs that are less creative. For simplicity, the experiments reported in Papers III IV are referred to in this dissertation as Experiment 1 and Experiment 2, respectively.
The experimentation paradigm used in this research is explanatory in nature since the main concern is to quantify the relationships and differences between the constructs associated with a specific software development activity—software designing. More generally, the goal of any explanatory research approach is to demonstrate the presence (or absence) of a causal relationship between the variables under consideration (Wohlin et al., 2012). In an experimentation paradigm, the factors of interest are already known and accounted for before the research is conducted. This is in direct contrast with an exploratory research approach, where new patterns emerge alongside as the research progresses. Here, the former research paradigm is also referred to as “quantitative research” and the latter as “qualitative research” (Creswell & Clark, 2017). However, it is sometimes advantageous to combine quantitative and qualitative research when investigating the phenomena of interest from different points of view (Driscoll, Appiah-Yeboah, Salib, & Rupert, 2007). This thesis employed an experimentation approach involving the categorical variables identified in the body of knowledge on cognitive biases, task structuring, and creativity (see Chapter 2) to explain and demonstrate the relationship between specification formality and design creativity.

The sections below describe the overall experimental design, followed by the variables, sampling, data collection and assessment technique, and statistical tests used to analyse the data.

**Overall experimental design**

This section provides a brief description of controlled experiments, together with the settings in which they were conducted and the rationale that drove the decision to use a controlled experiment research approach.

A controlled experiment “is an empirical inquiry that manipulates one factor or variable of the studied settings” (Wohlin et al., 2012, p.16). In a controlled experiment, the sample chosen from the target population is randomly divided into two or more groups, depending on the number of treatments. One group is usually given a placebo (i.e., either no meaningful treatment or a baseline treatment is applied), while all other groups are exposed to a specific treatment. Here, the applied treatment must be the only single factor that is varied within the studied settings. Any statistically significant differences between the groups in terms of the outcome of interest after a treatment is applied would most probably be due to the effect of the treatment rather than chance. Hence, one can infer that the treatment caused the observed variation. Controlled
experiments are usually conducted in a laboratory setting (or as close to it as possible), thus allowing greater control when randomly assigning participants to control and treatment groups and keeping all other variables constant.

This dissertation includes two between-subjects controlled experiments conducted in academic settings that made randomization of subjects possible. Both the studies had one treatment (i.e., desiderata framing) with two levels: requirements or prioritized requirements and ideas. These variables were based on the rationale mentioned in Chapter 2. In the context of this dissertation, one controlled experiment with one control and two treatment groups would have been an ideal research design. However, the study was broken into two separate experiments to achieve a sufficiently high statistical power, given the number of available participants. While Experiment 1 compared desiderata framed as requirements (treatment group) to desiderata framed as ideas (control group), Experiment 2 compared desiderata framed as prioritized requirements (treatment group) to desiderata framed as ideas (control group).

**Variables**

The experiments conducted as a part of this dissertation relied on several variables, categorized as independent variable and the dependent variables. These variables were conceptualized based on the research gaps (section 1.1) reported in this dissertation and operationalized based on the existing literature (Chapter 2). Table 3 provides a summary of the IVs and DVs used in this research.

A subjective assessment by expert judges was used to measure and quantify the DVs (i.e., originality and practicality) in both experiments, and both the DVs were measured on an ordinal scale.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Operationalized as</th>
<th>Type of Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desiderata Framing</td>
<td>Requirements</td>
<td>Independent</td>
</tr>
<tr>
<td></td>
<td>Prioritized Requirements</td>
<td>Independent</td>
</tr>
<tr>
<td></td>
<td>Ideas</td>
<td>Independent</td>
</tr>
<tr>
<td>Design Creativity</td>
<td>Design Originality</td>
<td>Dependent</td>
</tr>
<tr>
<td></td>
<td>Design Practicality</td>
<td>Dependent</td>
</tr>
</tbody>
</table>
Sampling

For both the experiments reported in Papers III and IV, the data were collected by soliciting participation from post-graduate students who were enrolled in computer science and software engineering courses.

For Experiment 1, the students were sampled by convenience among those enrolled in management and engineering courses at Lancaster University in the United Kingdom by using the relevant student mailing lists. A convenience sample of 19 females and 23 males with a mean age of 25 years (standard deviation 6.07) participated in this experiment. All participants received complementary lunch coupons for taking part in the study.

For Experiment 2, a convenience sample of 34 students (7 female and 27 male) from the University of Oulu in Finland was solicited. The average age of participants was 26 years (standard deviation 5.83). Although these participants did not receive any kind of compensation for participation, they were awarded extra credit as part of their coursework.

The participants in both experiments each signed a consent form to allow the collection and use of the data for the purpose of the experiments. None of the participants were informed, trained, or provided any instructions concerning the experimental task or the nature of the research itself. The identity of the participants remained completely anonymous throughout the research process, and in no way could the findings or the results of the experiments be traced back to reveal the identity of the participants.

Data collection and assessment technique

Each experiment was conducted in two parallel sessions, in separate rooms with similar dimensions and settings. Participants were randomly assigned to one of the two rooms and were given a list of desiderata framed as requirements, prioritized requirements, or ideas and several blank design templates. The participants were asked to complete a design task over 60 minutes, after which they were asked to fill out a post-task questionnaire including a manipulation check.

The data in the form of solution designs were de-identified, randomly shuffled, and given to two expert judges for assessment. Before the actual assessment, both judges discussed the meanings of “originality” and “practicality” and graded three random designs to establish a common understanding of the assessment procedure. The
judges then independently assessed the remaining designs. Reliability was evaluated by calculating the inter-rater agreement between the judges by using Cohen’s Kappa, a way of representing the standardized difference between two groups (Landis & Koch, 1977). Any disagreements were resolved by a third expert judge.

Data analysis and statistical tests

This section introduces the most important statistical tests used for processing the quantitative data collected from both experiments:

Descriptive statistics: Numerical descriptive statistics and their graphical representations provide a clear picture of how data are distributed. This information then helps researchers choose the most appropriate statistical technique to use for further data analysis. The main descriptive statistics used in this dissertation are as follows:

- Measures of a central tendency (e.g., mean, median, mode); this helps to measure a midpoint for the sample.
- Measures of dispersion (e.g., variance, standard deviation, variation interval) convey the sample’s level of variation from the central tendency and indicates how dispersed or concentrated the data are from the central tendency.

Null hypothesis significance testing (NHST): The role of hypothesis testing in controlled experiments is to check whether or not a given null hypothesis can be rejected. For instance, in Experiment 1, the variable investigated with hypothesis testing was design originality. The null hypothesis stated that requirements framing would not affect design originality, whereas the objective of the study was to show that requirements framing does indeed reduce design originality. The null hypothesis in both experiments had a level of significance of 0.05, which means that the null hypothesis could be rejected only when there was at least a 95% probability that it did not observe a false positive: rejecting the null hypothesis when in reality it was actually true (Fisher, 2006). Two statistical tests were used during the studies, according to the distribution of the DVs:

- A parametric test (independent samples t test) was used to compare the difference between the groups when the DVs were normally distributed.
- A non-parametric test (Mann-Whitney U test) was used to compare the difference between the groups when the DVs were not normally distributed.
Cross-Study Synthesis: A meta-analysis was conducted by combining the data from Experiments 1 and 2 into three groups—one control group (ideas framing) and two treatment groups (requirements framing and prioritized requirements framing). Again, the distributions of the DVs were compared across all three groups using the suitable NHST. Table 4 summarizes the different kinds of tests used in Paper III and Paper IV to analyse the data.

Table 4. Description of statistical tests used in controlled experiments.

<table>
<thead>
<tr>
<th>Statistical test</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Independent samples t</td>
<td>Parametric</td>
<td>To compare the differences between two groups when DVs are normally distributed</td>
</tr>
<tr>
<td>Mann-Whitney U</td>
<td>Non-parametric</td>
<td>To compare the differences between two groups when DVs are not normally distributed</td>
</tr>
<tr>
<td>Kruskal-Wallis</td>
<td>Non-parametric</td>
<td>An extension of Mann-Whitney U test for multiple samples</td>
</tr>
<tr>
<td>Levene’s non-parametric</td>
<td>NA</td>
<td>To test homogeneity of variance</td>
</tr>
<tr>
<td>Shapiro-Wilk</td>
<td>NA</td>
<td>To test normality of DVs</td>
</tr>
</tbody>
</table>

Correlations: Correlations are bivariate analysis techniques that measure the strength of association between two variables (e.g., originality and importance of specifications) and the direction of the relationship between the two. Pearson correlation is typically used when the data are normally distributed, and Spearman correlation is used when the data are not normally distributed or if the data are measured on an ordinal scale. The value of the coefficient ranges from +1 to -1. While a value of +1 indicates a complete association between the variables, -1 indicate no association. A ‘+’ sign indicates a positive relationship, and a ‘-’ sign indicates a negative relationship between the variables.

Effect size: Effect size is a collection of statistical technique used to measure the magnitude of the effect of a treatment (Ellis, 2010). Effect size usually evaluates the practical significance of an experiment’s results, in contrast to the p-value, which evaluates only the experiment’s statistical significance. The specific effect size used in both experiments was Cliff’s ($\delta$), which is calculated as: $\#(X_1 > X_2) - \#(X_1 < X_2)/n_1n_2$, where $X_1$ and $X_2$ are the scores within the two groups, and $n_1$ and $n_2$ are the sizes of the sample groups. Here, the cardinality symbol # indicates the count (Macbeth, Razumiejczyk, & Ledesma, 2011). The effect size is usually reported alongside a confidence interval (CI). The CI is the range of values in which the effect size can vary.
A CI is usually reported at 95% precision, meaning that is only a 5% possibility that the effect size will not fall within the CI values (Fritz, Morris, & Richler, 2012).

### 3.2.4 Protocol study

RQ2.2 was answered by conducting a *dialog-based* protocol study, as reported in Paper V. A protocol study is a type of exploratory research approach that provides a technique to uncover the thought process while performing a task or solving a problem. In a protocol study, subjects give verbal reports (i.e., they “think aloud”) on the cognitive processes that they are experiencing during a task. The underlying principle of this technique is “that any verbalization produced by a subject while solving a problem will directly represent the contents of the subject’s working memory” (Ericsson & Simon, 1993). While alternative techniques such as interviews or post-study questionnaires provide an incomplete report due to forgetfulness or selective reporting of thoughts (Ericsson & Simon, 1993), concurrent verbalization of thoughts can help to obtain real-time insights into the knowledge used by a subject while solving a problem or completing a task (Hughes & Parkes, 2003). While a complete review of the applications of protocol analysis is beyond the scope of this dissertation, this technique has been extensively used in psychology (Ericsson & Simon, 1993), medicine (Hashem, Chi, & Friedman, 2003), and engineering design (Dorst & Dijkhuis, 1995). The three main types of protocols are introspective, retrospective, and think-aloud (Gray & Anderson, 1987). While the introspective and think-aloud approaches involves subjects reporting their thoughts during the task, retrospective protocol reports are collected once the task is completed (Ericsson & Simon, 1993).

Specifically in SE, think-aloud protocol analysis is commonly used to analyse human behaviour in processes like pair programming (Jain, Muro, & Mohan, 2006), information processing in software development (Mohan, Kumar, & Benbunan-Fich, 2009), and software development in general (Hale, Sharpe, & Hale, 1999). Broadly speaking, a think-aloud protocol approach helps not only to analyse what the individual subjects are doing (i.e., task completion) but also why (i.e., the rationale underlying a decision). For example, concurrent verbalization made while creating conceptual designs could uncover the underlying cognitive processes and decision models used by the designers while they actively make sense of the problem and conceive solutions. However, think-aloud protocol analysis suffers from two major issues, as summarized by Xu and Rajlich (2005):

52
1. Placebo effect, where subjects may deliberately try to verbalize the desired data (Rosenthal, 1976).
2. Hawthorne effect, which explains a change in the subject’s behaviour in the presence of a facilitator or researcher (Adair, 1984).

To overcome these inherent issues, a dialog-based protocol study (Xu & Rajlich, 2005) was used. A dialog-based protocol is an alternative to a think-aloud protocol; in the latter, only one subject speaks, while in the dialog-based protocol, the conversation between two subjects is recorded. Here, the dialog or conversation between two subjects while completing a task results in more rich and detailed data and significantly reduces the Hawthorne and placebo effects. This technique is analogous to pair programming (Williams, Kessler, Cunningham, & Jeffries, 2000), in which two programmers co-operate, discuss, and verbalize their thoughts to enhance both parties’ understanding.

**Sampling**

A convenience sample of 18 software developers from Company X and 24 post-graduate students enrolled in the information processing science program at the University of Oulu in Finland were recruited for this study. The author paired participants based on their availability. The software developers did not receive any compensation for participating; the students received extra credit as a part of their coursework. Each participant signed a consent form to record the proceedings and use the data for analysis and reporting.

**Study design and data collection**

The data collection process was completed between September and December 2016 on the premises of the company and the university. Each pair was scheduled for a session in a quiet setting where they were asked to complete a design task from a list of requirements statements; while doing so, they were asked to discuss, co-operate, and verbalize their thought process. Before starting each session, the participants were again reminded to keep thinking aloud. However, the participants were prompted even during the task if they spoke in a language other than English or made any design moves without thinking aloud or discussing them with each other. There were no other interactions between the author and the participants, to avoid any significant interference.
with participants’ thought processes. The dialogues between the participants were audio recorded.

**Data analysis**

The recorded dialogues were transcribed and cleaned by removing all the verbal static (e.g., “um”, “ah”, “uh”). The data were then stored in NVIVO software for qualitative analysis. The data were analysed in two phases—1) inductive process coding and 2) closed coding based on an a priori coding scheme.

*Process coding* involved analysing the data by using inductive coding (Saldaña, 2015). Each transcript was coded line by line using words ending with “-ing”, where each assigned code reflected the action contained in dialogues that shared similar characteristics. Over many iterations, some of the codes were re-worded, subsumed by other similar codes, or dropped altogether. All codes that conveyed a particular process or action were categorized together to form themes (Cruzes & Dyba, 2011).

In contrast, *closed coding* involved re-coding the data using a previously developed coding scheme that was based on the theoretical conceptualization of fixation (see Chapter 2). This was done to compare the instances of fixation against the instances of more critical thinking. Finally, the number of instances of each category were counted and compared to show the levels of support for fixation and critical thinking.

### 3.3 Summary of the research process

Research conducted as part of phase 1 involved aggregating and summarizing the available research involving cognitive biases and creativity in SE to identify the relevant scientific literature concerning fixation and framing effect and to understand the way in which the categorical variables used in this dissertation are conceptualized in the research literature (e.g., different levels of specification formality and creativity as a cognitive process that produces original and practical artifacts). This knowledge was used to formulate the hypotheses (i.e., *more formal and structured desiderata framing will produce less creative designs*), which were later tested through controlled experiments conducted in phase 2. The research methods involved in this phase were an SMS and interviews. While the SMS provided a high-level overview of the topic under investigation the exploratory interview study helped to investigate practitioners’ perspectives on the role of creativity in SE.
In phase 2, controlled experiments were used to investigate the impact of framing desiderata as ideas, requirements, or prioritized requirements on design creativity, where creativity was evaluated in terms of design originality and practicality. This explanatory research approach helped to explain the differences in the originality and practicality scores due to variation in specification formality.

However, merely quantifying the differences and demonstrating a causal relationship do not provide a clear understanding of the phenomena under investigation. Therefore, in phase 3, an exploratory research approach in the form of a dialog-based protocol study was also conducted to understand the reasons for this difference. This research method helped to explore the cognitive mechanisms that may explain the differences in creativity levels due to specification formality. The evidence gathered from the explanatory and exploratory research approaches was interpreted together to generate a theoretical framework explaining the effect of specification formality on design creativity.
4 Original research publications

This section presents the publications included in the dissertation. All are ranked in the Finnish Publication Forum (JUFO, from the Finnish “Julkaisufoorumi”)\(^3\), which assesses the quality of academic research. Papers I-IV have been published in peer-reviewed international conference proceedings and journals as follows: Paper I: *IEEE Transactions on Software Engineering* (IEEE TSE 2018; JUFO level 3); Paper II: the *Forty-third Euromicro Conference on Software Engineering and Advanced Applications* (SEAA 2017; JUFO level 1); Paper III: the *Thirty-sixth International Conference on Software Engineering* (ICSE 2014; JUFO level 2); Paper IV: *IEEE Transactions on Software Engineering* (IEEE TSE 2019; JUFO level 3); Paper V is currently under review at the *Forty-second International Conference on Software Engineering* (JUFO level 2). Table 5 shows how each paper addresses various research questions. The following subsections provide more details on the motivations and findings of each paper and the ways in which each fits into the dissertation.

The papers summarized below follow the same basic structure—first, we introduce the rationale that motivated the study, followed by a brief description of the study and the salient results. Finally, the publications’ main contributions are reported in the context of this dissertation. A more detailed account of the respective findings is reported in Chapter 5. The respective publishers granted permission to republish Figures 3–6 and Tables 6–9.

\(^3\)http://www.julkaisufoorumi.fi/en
Table 5. Summary of the contribution of each publication presented in this dissertation.

<table>
<thead>
<tr>
<th>Paper ID</th>
<th>Purpose</th>
<th>Main Findings</th>
<th>Sub-RQs</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Systematically map, aggregate, summarize, and synthesize the available literature on cognitive biases in SE to understand the state of the art research and provide basis for future investigations.</td>
<td>A: The study identified 65 primary studies investigating 37 cognitive biases. Most of the research is focused on investigating the antecedents and effects of biases. While the most frequently investigated cognitive biases belong to interest and stability bias, social and decision biases are least investigated. Significant confusion exists concerning the interpretation and theoretical conceptualization of these biases.</td>
<td>RQ1.1</td>
</tr>
<tr>
<td>II</td>
<td>Explore and understand the way in which the SE research literature and practitioners conceptualize, improve, and assess creativity.</td>
<td>B: SE researchers and practitioners appear to perceive creativity differently, with minimal consensus, which hinders the development of evidence-based techniques and approaches to measure and enhance creativity.</td>
<td>RQ1.2</td>
</tr>
<tr>
<td>III</td>
<td>A between-subjects randomized controlled trial to investigate the effects of framing desiderata as requirements on design creativity.</td>
<td>C: Framing desiderata as requirements significantly reduced design creativity.</td>
<td>RQ2.1</td>
</tr>
<tr>
<td>IV</td>
<td>A between-subjects randomized controlled trial to investigate the effects of framing desiderata as requirements and prioritized requirements on design creativity.</td>
<td>D: Framing desiderata as prioritized requirements significantly reduced design creativity.</td>
<td>RQ2.1</td>
</tr>
<tr>
<td>V</td>
<td>Explore the cognitive mechanisms that may explain why framing desiderata more formally as requirements reduces creativity during software design.</td>
<td>E: Participants were found to uncritically accept aspects of the task (e.g., requirements, priority levels), adopt features of known examples without any critical reflection, and reject divergent ideas.</td>
<td>RQ2.2</td>
</tr>
</tbody>
</table>
4.1 **Author’s contributions**

The author had significant involvement in every phase of the research and is the leading author of each publication.

For Paper I, the author was involved in all steps of conducting SMS research, including the literature search, data extraction, analysis, and reporting the findings in the paper. The author led the writing of the paper, addressed the reviewers’ comments, and prepared the final draft, which was reviewed by the entire author team. The second author in Paper I contributed equally to each step, from inception to acceptance.

The author conceived and designed both the empirical studies reported in Paper II. With the exception of executing the interviews, the author was involved in all steps of executing the SMS and interviews, including data analyses, data triangulation, and reporting the findings. The second author was responsible for conducting the interviews, transcribing the recorded data, and contributed equally to data analysis and triangulation. In addition, the author addressed the reviewers’ comments and wrote all the sections of the paper, which were reviewed by the co-authors.

The author conceived, designed, collected, and analysed the data for the empirical studies reported in Papers III, IV, and V. The author led the writing of the papers, prepared the figures and tables, addressed the reviewers’ comments, and was responsible for preparing the final drafts of all the papers. The entire author teams reviewed all three papers.

4.2 **Paper I - Cognitive biases in software engineering: A systematic mapping study**

**Rationale:** The first paper of this dissertation provides a comprehensive view of research on cognitive biases in SE. The study helped to understand the state-of-the-art research by focusing on bias antecedents, effects, and mitigation techniques and provided guidelines for future research investigations concerning fixation and framing effect.

**Description:** An SMS was conducted to aggregate, summarize, and categorize knowledge involving cognitive biases in SE research that is dispersed across the primary studies, based on established guidelines (Kitchenham, 2010; Kitchenham et al., 2010; Petersen et al., 2008, 2015)).

**Results:** The study identified 65 primary studies investigating 37 cognitive biases. The most investigated categories are *interest biases* (e.g., confirmation bias, valence
effect), followed by stability biases (e.g., anchoring and adjustment bias, status-quo), whereas the categories of social biases (e.g., bandwagon effect) and decision biases (e.g., hyperbolic discounting, sunk-cost) are least investigated. While Figure 3 summarizes the frequency of occurrence of each cognitive bias broken down by SE knowledge area, Table 6 provides the number of primary studies investigating cognitive biases in each category. Besides providing a road-map and recommendations for future investigations, the major findings from this SMS are summarized as follows:

- Most SE research is focused on investigating the causes and effects of multiple cognitive biases. However, the psychological and sociological factors underlying the cognitive biases remain under-investigated.
- The most common research method for investigating cognitive biases was controlled experiments; most of the empirical research focused on establishing causality rather than exploring facts or developing theories. However, no replications, families of experiments, or meta-analyses were found.
- Most of the primary studies were disconnected and did not refer to other studies concerning the same bias. Moreover, empirical studies investigating cognitive biases conceptualized and defined biases inconsistently and even used biases with similar underlying cognitive mechanisms interchangeably.

Contributions: In the context of this dissertation, this paper addressed RQ1.1 by providing a comprehensive view of the state-of-the-art research concerning cognitive biases (including fixation and framing effect) in SE and established a basis for setting up the subsequent investigations. Specifically, this study motivated future research to focus on under-investigated biases, develop effective bias mitigation techniques, and develop new theories that can explain the manifestation of cognitive biases in SE. Additionally, this paper provided several useful insights about the conceptualization of fixation and framing effect in SE, which helped the researcher to choose the appropriate type of research method to address the research gaps (see section 1.1) in Papers III, IV and V.

4.3 Paper II - Perceptions of creativity in software engineering research and practice

Rationale: Assessing creative performance is challenging because creativity is multidimensional, difficult to quantify, subjective, and not clearly understood. In this context, building consensus around dimensions of creativity is critical to developing appropriate
<table>
<thead>
<tr>
<th>Cognitive Bias</th>
<th>Computing foundations</th>
<th>Configuration</th>
<th>Design</th>
<th>Economics</th>
<th>General</th>
<th>Management</th>
<th>Mathematical foundations</th>
<th>Models and Methods</th>
<th>Process</th>
<th>Quality</th>
<th>Requirements</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchoring/adjustment</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>Attention</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Availability</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Bandwagon effect</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Belief perseverance</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>(Over)-confidence</td>
<td>1</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>Confirmation</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>Contrast effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Default</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Endowment effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Fixation</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Framing effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Framing effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Halo effect</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Hindsight</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Hyperbolic discounting</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>IKEA effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Impact</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Information</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Invincibility</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Mere exposure effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Misery info. processing</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Misleading information</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Neglect of probability</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Normalcy effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>(Over)-optimism</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Primacy/recency</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Representativeness</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Selective perception</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Semantic fallacy</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Semmelweis reflect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Status quo</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Sunk cost</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Time-based</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Valence effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Validity effect</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Wishful thinking</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

Total 0 0 22 39 0 0 5 2 44 0 2 2 0 4 15 7

Fig. 3. Paper 1 results: Cognitive biases investigated in SE knowledge areas, ©2018 IEEE.
Table 6. Paper 1 Results: Primary studies organized into Fleischmann et al., (2014) taxonomy, ©2018 IEEE.

<table>
<thead>
<tr>
<th>Bias category</th>
<th>Cognitive biases</th>
<th># of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interest</td>
<td>Confirmation, IKEA effect, valence effect, validity effect, wishful thinking</td>
<td>30</td>
</tr>
<tr>
<td>Stability</td>
<td>Anchoring and adjustment, belief perseverance, default, endowment effect, status quo</td>
<td>29</td>
</tr>
<tr>
<td>Action-oriented</td>
<td>(Over)-confidence, impact, invincibility, miserly information processing, misleading information, normalcy effect, (over)-optimism</td>
<td>22</td>
</tr>
<tr>
<td>Pattern recognition</td>
<td>Availability, mere exposure effect, fixation, Semmelweis reflex</td>
<td>19</td>
</tr>
<tr>
<td>Perception</td>
<td>Attentional, contrast effect, framing effect, halo effect, primacy and recency effect, selective perception, semantic fallacy</td>
<td>9</td>
</tr>
<tr>
<td>Memory</td>
<td>Hindsight, time-based bias</td>
<td>6</td>
</tr>
<tr>
<td>Decision</td>
<td>Hyperbolic discounting, infrastructure, neglect of probability, sunk cost</td>
<td>3</td>
</tr>
<tr>
<td>Social</td>
<td>Bandwagon effect</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>Representativeness, information</td>
<td>9</td>
</tr>
</tbody>
</table>

techniques and practices to assess and ultimately improve creativity (T. M. Amabile, 1982). Moreover, the role of creativity in SE can be better understood only by considering the perspectives, expectations, and reservations of SE researchers and practitioners alike. Consequently, this paper reports an exploratory study that investigates how creativity is understood, improved, and assessed in SE research and practice.

Description: This paper reports a multi-methodological study (Morse, 2003) comprising an SMS followed by a constructivist interview study. The systematic mapping approach helped to aggregate and synthesize the available evidence concerning how SE research literature characterizes creativity. The SMS protocol was compiled based on established guidelines (Kitchenham, 2010; Petersen et al., 2015). This was followed by a cross-sectional interview study to investigate practitioners’ perspectives on creativity. Seventeen SE practitioners from five countries were interviewed using a semi-structured interview guide that was developed by following (DiCicco-Bloom & Crabtree, 2006; Whiting, 2008). Finally, data triangulation was conducted, with the findings from the
mapping and qualitative analysis compared and contrasted to reveal the perceptual differences concerning the way creativity is conceptualized in SE research and practice. **Results:** Eighty-four primary studies and qualitative analysis of seventeen interviewees support our assumption that creativity plays a critical role in SE processes. The subsequent analysis reveals very few cases of consensus between the two communities, such as creativity as a cognitive process that produces original and useful ideas. However, the results also reveal major disagreements and confusion between researchers and practitioners, especially in how creativity is assessed and improved. More generally, both researchers and practitioners seem to adopt very constrained views of creativity (e.g., conceptualizing creativity in terms of a creative product), often disagreeing and disregarding other perspectives, whether of person, place, persuasion, process, or potential. While most interviewees were oblivious to the state-of-the-art techniques and practices that could enhance or assess creative performance, there seems to be no consensus in SE research concerning any approaches or techniques to objectively assess creativity.

This confusion hinders the development of a comprehensive understanding of the way in which creativity is conceptualized in SE, which in turn impedes any attempt to develop evidence-based and scientifically valid creativity assessment techniques. **Contributions:** This paper addresses RQ1.2 and drives the empirical studies reported in Papers III and IV by making the following major contributions to this dissertation:

- Creativity as a multidimensional construct, where a creative artefact should not only be original but also practically useful.
- There exists no consensus in SE research on an objective metric to assess creativity. Hence, the most favoured approach to evaluate creativity of a product is a subjective assessment by multiple domain experts, and based on perceived novelty or usefulness of an artefact.

### 4.4 Paper III - Requirements fixation

**Rationale:** This paper reports a controlled experiment to investigate the impact of framing desiderata as requirements on design creativity. Specifically, this paper tests the hypothesis that framing desiderata as structured and formal requirements results in less creative designs.

**Description:** A between-subjects randomized controlled trial was conducted using a convenience sample of 42 post-graduate students. Participants were randomly assigned
to either a treatment or a control group. All participants were given the same set of desiderata and asked to create conceptual designs of a mobile application. However, the treatment group received desiderata framed as requirements, whereas the control group received desiderata framed as ideas. The ideas framing was chosen as presenting desiderata because “ideas” is more ambiguous and uncertain than presenting desiderata as “requirements” and to minimize the difference between the two groups. The participants were also asked to rate, on a five-point scale, the importance of specifications for guiding their designs. Two expert judges independently assessed creativity in terms of the originality of the conceptual designs on a five-point ordinal scale from low (1) to high (5). A third expert judge resolved any disagreements.

**Results:** The results of the experiment showed a significant difference in the originality grades between the groups. Specifically, participants who received ideas framing produced significantly more original designs than participants who received requirements framing (Mann-Whitney U test; U=116.5, p=0.004) with an effect size (Cliff’s $\delta$= -0.49) indicating a “high” effect with 95% CI [-0.74, -0.13] (Kitchenham et al., 2017). Figure 4 shows differences in the originality grades between the two groups. Moreover, the treatment group (Median=4) rated the specifications as more important than the control group (Median=3).

**Contributions:** In the context of this dissertation, this paper addresses RQ2.1 and suggests that presenting desiderata as “requirements” significantly reduces design creativity. Specifically, this paper hypothesizes that presenting ostensible desiderata as formal requirements leads designers to fixate on desiderata, resulting in solution designs that are less creative.

### 4.5 Paper IV - Requirements framing affects design creativity

**Rationale:** This paper investigates the effect of framing desiderata as requirements and prioritized requirements on design creativity. Specifically, this paper extends the idea from Paper III in two ways—first, the data from the previous experiment (Experiment 1) were re-analysed to measure the effect of the treatment on the practicality of design concepts; second, an additional experiment was conducted to investigate the effect of framing desiderata as prioritized requirements on the originality and practicality of solution designs. This study tests the hypothesis that more structured desiderata would lead to less original, more practical designs.
Description: Experiment 2 was conducted using a convenience sample of 34 post-graduate students and followed the same basic design as Experiment 1. However, in Experiment 2, the treatment group received desiderata framed as prioritized requirements. Prioritizing the requirements adds more structure and clarity to the desiderata than the simple notion of requirements. Two expert judges independently assessed creativity in terms of the originality and practicality of the designs on a five-point ordinal scale from low (1) to high (5). A third expert judge resolved any disagreements. As a post-experimental manipulation check, the participants were asked to rate, on a five-point

Table 7. Statistical differences in originality grades.

<table>
<thead>
<tr>
<th>Statistical test</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U test</td>
<td>U=116.5, p=0.004</td>
<td>U=84, p=0.02</td>
</tr>
<tr>
<td>Effect size (Cliff’s δ)</td>
<td>-0.49</td>
<td>-0.43</td>
</tr>
<tr>
<td>95% CI</td>
<td>[-0.74, -0.13]</td>
<td>[-0.71, -0.05]</td>
</tr>
</tbody>
</table>
Table 8. Statistical differences in practicality grades.

<table>
<thead>
<tr>
<th>Statistical test</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mann-Whitney U test</td>
<td>U=128.5, p=0.018</td>
<td>U=77, p=0.02</td>
</tr>
<tr>
<td>Effect size (Cliff’s δ)</td>
<td>0.41</td>
<td>0.46</td>
</tr>
<tr>
<td>95% CI</td>
<td>[0.06, 0.67]</td>
<td>[0.07, 0.73]</td>
</tr>
</tbody>
</table>

scale, the importance of the list of specifications in guiding their designs and how confident they felt about having their solution designs be implemented. A meta-analysis was conducted, in which data from both experiments were combined to check whether there were any differences among the three groups: control (ideas framing), treatment-1 (requirements framing), and treatment-2 (prioritized requirements framing).

**Results:** The results showed significant differences in the originality and practicality grades between the groups. Participants who received desiderata framed as prioritized requirements produced less original designs than participants who received ideas framing (Table 7). In contrast, participants who received desiderata framed as requirements or prioritized requirements created significantly more practical designs than participants who received ideas framing (Table 8). The differences in the originality and practicality grades for both the experiments are shown in Figures 5 and 6, respectively. Like the experiment reported in Paper III, participants in the treatment group (Median=4) rated the specifications as more important than the control group (Median=2). Participants in the treatment group (Median=4) also felt more confident that their designs could be practically implemented than participants in the control group (Median=3).

The results of the cross-study synthesis (Table 9) highlight the differences between the three groups and show that providing more structure to desiderata leads to less original designs (Spearman correlation, rho=0.430, p<0.001). Moreover, both treatment groups exhibit better design practicality than the control group (Kruskal-Wallis H=11.978, p=0.003). Further analysis showed that fixation mediated the relationship between framing and practicality (importance of desiderata is positively correlated with practicality—rho=0.297, p=0.005) but not originality (importance of desiderata is negatively correlated with originality—rho=-0.144, p=0.108). However, the latter was not statistically significant.

**Contributions:** This paper extends and builds on the results of Paper III. Both experiments show that presenting desiderata as formal requirements statements affects design creativity. Specifically, the results suggest that framing ostensible desiderata as requirements or prioritized requirements results in less original but more practical
Fig. 5. Frequency of originality and practicality grades (Experiment 1), ©2019 IEEE.
Fig. 6. Frequency of originality and practicality grades (Experiment 2), ©2019 IEEE.
designs. The experiments presented in Paper III and Paper IV together answer RQ2.1 and establish a negative relationship between specification formality and design creativity, as illustrated in Figure 7.

![Figure 7: Fixation mediates the relationship between problem structuring and creativity, ©2019 IEEE.](image)

Note: Unfilled arrows indicate causation; unfilled diamond indicates aggregation; plus and minus signs indicate magnitude of effect.

**Table 9. Cross-study synthesis, ©2019 IEEE.**

<table>
<thead>
<tr>
<th></th>
<th>Control (Ideas)</th>
<th>Treatment 1 (Requirements)</th>
<th>Treatment 2 (Prior. req.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean originality</td>
<td>3.26</td>
<td>2.67</td>
<td>2.06</td>
</tr>
<tr>
<td>Mean practicality</td>
<td>2.26</td>
<td>3.48</td>
<td>3.12</td>
</tr>
<tr>
<td>Number (n)</td>
<td>38</td>
<td>21</td>
<td>17</td>
</tr>
</tbody>
</table>

### 4.6 Paper V - Requirements fixation: A dialog-based protocol study

**Rationale:** The experiments reported in Papers III and IV showed that a more formal and structured presentation of desiderata reduced creativity. The experiments hypothesized that presenting desiderata as formal requirements triggered fixation, which reduced the creativity of solution designs. However, randomized controlled trials like those reported in Papers III and IV are inappropriate for uncovering and explaining the cognitive mechanisms underlying these causal effects. This empirical study, therefore explores the cognitive mechanisms that may explain why framing desiderata as formal requirements reduces software design creativity.
**Description:** A dialog-based protocol analysis (Xu & Rajlich, 2005) was conducted with 21 pairs comprised of software practitioners and post-graduate students. The rationale for using protocol analysis is that any verbalization made by participants while solving a problem represents the cognitive functioning of participants’ working memory. The conversations between the participants were recorded, transcribed, and qualitatively analysed in two phases. First, data were analysed by coding the transcripts inductively, followed by closed coding, where data were re-coded based on a priory coding scheme. The number of instances of each code showed the support for each category. Based on the results, a theory explaining the relationship between specification formality and design creativity was proposed.

**Results:** The analysis shows that design actions associated with fixation were more frequent than design actions associated with critical thinking. The exploratory results suggest that participants tended to uncritically accept requirements and priority levels and adopt features of known examples without any critical reflection, avoiding divergent ideas. More specifically, presenting more structured or over-rationalized desiderata—mainly as requirements—triggers fixation and impedes critical thinking.

**Contributions:** This paper answers RQ2.2 and elaborates on the concept of specifications formality reducing design creativity. While the experiments reported in Papers III and IV established causality between specification formality and creativity, this paper contributes by exploring the cognitive mechanisms that explain this relationship. The results of this paper suggest that the negative relationship between specification formality and creativity is mediated by fixation and critical thinking. Presenting desiderata more formally as requirements statements appears to induce requirements fixation and impedes critical thinking.
5 Results, discussions, and limitations

This chapter reports a synthesis of the findings from the empirical studies presented in Chapter 4. Figure 8 presents the five main findings of this dissertation, labelled A to E. Sections 5.1 and 5.2 answer sub-research questions RQ1 and RQ2, respectively, and interpret the findings in the context of the dissertation. After a summary is presented in section 5.3, limitations and threats to the validity of this research appear in section 5.4.

5.1 Answer to RQ1: What is the current state of research concerning cognitive biases and creativity in SE?

This research question was partially answered by systematically reviewing and summarizing the existing studies concerning cognitive biases in SE research. An SMS was conducted to provide a comprehensive view of research involving cognitive biases in SE research and to identify the way in which cognitive biases such as fixation and framing effects manifest themselves in myriad SE activities. SE research involving cognitive biases showed the following five notable characteristics:

1. SE research often investigates multiple causes and effects of biases.
2. SE research rarely investigates techniques for preventing biases or mitigating their effects.
3. The psychological and sociological mechanisms underlying many cognitive biases are poorly understood.
4. SE research on cognitive biases is highly fragmented and rarely replicated.
5. Biases are inconsistently defined across studies, and the relationships between biases are not well understood.

RQ1 was also partially answered by exploring the current state of research concerning the role of creativity in SE. A multi-methodological approach (Morse, 2003) comprised of an SMS, and a subsequent constructivist interview study was conducted to investigate how creativity is understood, assessed, and improved in SE research and practice. Analysing 84 primary studies and 17 interviews showed that creativity is an important topic of interest in SE research on creativity, which showed the following four significant characteristics:
Fig. 8. Summary of the main findings of the dissertation.

**Research Gaps**

1. Lack of a comprehensive body of knowledge that summarizes the available literature in SE research involving cognitive biases, particularly concerning fixation and framing effect.
2. Lack of shared agreement between SE research and practice concerning the way in which creativity is understood and assessed in SE.
3. The question of whether or not more structured desiderata framing (i.e. specification formality) impedes design creativity.

**RQ1.1 - What is the current state-of-the-art in SE research involving cognitive biases?**

*Finding A:* A constant and sustained interest in bias-related research is clear with scarcity of empirical research on bias-mitigation approaches and lack of established theories to explain and interpret the cognitive biases.

**RQ1.2 - How is creativity conceptualized, assessed and improved in SE?**

*Finding B:* SE researchers and practitioners appear to perceive creativity differently with very little consensus, which hinders development of evidence-based techniques and approaches to measure and enhance creativity.

**RQ2.1 - Does the framing of desiderata affect design creativity?**

*Findings C & D:* Participants who received desiderata framed as requirements or prioritized requirements created significantly less creative designs than participants who received desiderata framed as ideas.

**RQ2.2 - What cognitive behaviors can explain the relationship between specification formality and design creativity?**

*Finding E:* Participants were found to uncritically accept aspects of the task (i.e. requirements, priority levels), adopt features of known examples without any critical reflection and rejected divergent ideas.
1. Both SE research and practice generally conceptualize creativity as a cognitive process that produces not only novel but also useful ideas or artifacts.

2. However, practitioners were found to conflate this notion of creativity with other quality attributes such as simplicity, ease of use, marketability, and so on.

3. While the most popular creativity-inducing and -enhancing technique is brainstorming, the most common creativity assessment technique is subjective assessment of creative products by multiple experts.

4. The analysis, however, showed a high degree of confusion between the two communities, which undermines the role of creativity in SE.

5.1.1 Answer to RQ1.1: What is the current state of the art in SE research involving cognitive biases?

Paper I was conceived to answer RQ1.1. It identified 65 primary studies exploring 37 cognitive biases, which were further classified by using an existing taxonomy by Fleischmann et al. (2014) to obtain a higher-level view of the kinds of cognitive biases that are more frequently investigated. While the most commonly investigated cognitive biases belonged to interest and stability biases, the categories of social and decision biases were the least investigated. Meanwhile, software project management was the most investigated SE activity for cognitive biases, followed by software design and software development. The current state-of-the-art SE research involving cognitive biases is illuminated by the following summary of the main findings:

- Most of the existing SE research is focused on investigating multiple antecedents and effects of cognitive biases. However, the psychological and sociological mechanisms underlying many cognitive biases are poorly understood.
- Debiasing approaches for only 6 (of 37) cognitive biases were found in primary studies, none of which were empirically evaluated; all the techniques discussed in primary studies were untested suggestions. This shows that SE research rarely investigates techniques for preventing biases or mitigating their effects.
- Controlled experiments were the most preferred research method for investigating cognitive biases in SE. However, no replications, families of experiments, or meta-analyses were found. Therefore, SE research on cognitive biases is highly fragmented and rarely replicated. It is nevertheless critical to understand not only whether but also how cognitive biases affect productivity, quality, creativity, and SE success. Moreover,
a better understanding of the mechanisms through which biases affect SE success could help develop effective bias mitigation techniques.

- Most primary studies are disconnected. That is, many investigations were conducted in complete isolation without any reference to previous SE research on the same bias or conceptualized them differently, resulting in a great deal of confusion among the studies investigating the biases.

In the context of this dissertation, the answer to RQ1.1 helped to gain a measure of the interest in bias-related investigations in SE, consolidated the scientific literature required to position further investigations concerning fixation and framing effects, and identified the sub-topics concerning cognitive biases and SE activities that warrant more empirical investigations.

5.1.2 Answer to RQ1.2: How is creativity conceptualized, assessed, and improved in SE?

To answer RQ1.2, Paper II combined an SMS of SE research literature with an interview study of practitioners. The subsequent data triangulation (i.e., comparing and contrasting the dimensions of creativity that emerged from the mapping study and qualitative analysis) revealed some agreement but also some major differences in the way creativity is conceptualized, assessed, and improved in SE.

Both SE research and practice conceptualize creativity as a cognitive process that produces novel and useful ideas or artifacts—creative products. However, practitioners appear unconcerned with either or both of these dimensions. Instead, they conflate creativity with more general notions of dimensions of quality like simplicity, ease of use, marketability, and so on. This shows that, while a consensus exists among SE researchers about conceptualizing creativity in terms of the originality and practicality of artefacts, there remains a degree of confusion among SE practitioners.

Moreover, both researchers and practitioners appear to ignore or even refuse to acknowledge creativity in terms of a process, a personality trait, potential, or the context (place) of the work. This confusion impedes the development of a comprehensive understanding of the concept of creativity in SE, which in turn hinders scientifically valid attempts to develop objective approaches to assess creativity (Chi, 2006; Dillon, 1982). Hence, subjective assessment of creative products by experts or peers is the most common creativity assessment technique in SE (Adams, 2005; Baer & Kaufman, 2005).
Finally, brainstorming and sketching are the only two creativity-inducing and
-enhancing techniques that were mentioned by both primary studies and the interviewees.
However, as many as 16 practitioners considered the “freedom they enjoyed at work” as
a factor that influenced their creative practices. This suggests that these practitioners
were not aware of the state-of-the-art practices to enhance creativity that had been
proposed in SE literature such as creative problem solving (DeFranco-Tommarello,
Hiltz, Deek, Perez, & Keenan, 2003; Maiden, Gizikis, & Robertson, 2004), unfamiliar
connections (Bhowmik, Niu, Mahmoud, & Savolainen, 2014; J. Cao, Fleming, &
Burnett, 2011), and conducting creativity-enhancing workshops (Horkoff, Maiden, &
Lockerbie, 2015; Maiden, Manning, Robertson, & Greenwood, 2004).

To summarize, both SE research and practitioners appear to have a notably limited
understanding of creativity. Dimensions of creativity, other than product creativity,
remain under-investigated in SE research literature. This is highly problematic, as only
building a consensus around the dimensions of creativity can inform evidence-based
practices and techniques to assess and improve creativity in SE.

5.2 Answer to RQ2: How does specification formality affect design
creativity?

The answer to RQ2 is that more formal and structured framing desiderata significantly
reduces creativity of designs, where the negative relationship between the formality of
the desiderata presentation and design creativity is mediated by fixation.

Papers III and IV report two experimental studies that investigated the effects of
a more formal and structured framing of desiderata on design creativity. The results
of both experiments show that framing desiderata as “requirements” or “prioritized
requirements” resulted in less original but more practical designs, significantly affecting
design creativity. These findings are analogous to previous research conducted on design
fixation that investigated how providing designers with explicit examples reduces their
creativity (Finke, 1996; Zahner et al., 2010). This dissertation extends the stream of this
idea by demonstrating that framing desiderata as requirements similarly impedes design
creativity, even without providing any explicit examples.

More generally, designers tend to fixate on given desiderata when framed as
more structured, clear, and rationalized requirements, affecting their ability to think
critically. The results of the two experiments can be related to mental-set fixation—“a
situation where individuals restrict the use of their abilities due to situationally induced
bias” (Jansson & Smith, 1991). Specifically, the results highlight the innate tension between design creativity and system requirements. The findings reveal that the way a specification is framed might be just as important as its contents.

However, conducting experiments to establish the relationship between the specification formality and design creativity is not enough to fully explore the cognitive mechanisms underlying these causal effects. In such a scenario designers fixating on desiderata framed as requirements, can be considered only one of the plausible explanations. Consequently, a dialog-based protocol study was conducted to explore the mechanisms that might explain why framing desiderata as requirements reduces software design creativity. The results of a subsequent qualitative analysis reported in Paper V reveal that design actions associated with fixation were much more frequent than design actions associated with critical thinking. More generally, designers tended not to critically evaluate the structure of a given task when it was presented in a more formal way.

The combined results from the experiments and protocol study (findings C, D, and E) suggest that the negative relationship between specification formality and design creativity is mediated by fixation and critical thinking. However, designers tend to fixate much more when desiderata are framed as formal requirements, which impedes design creativity.

5.2.1 Answer to RQ2.1: Does the framing of desiderata affect design creativity?

To answer this research question, two experiments (Experiments 1 and 2) were conducted to investigate the impact of presenting desiderata as requirements and prioritized requirements on design creativity. We continue this section by discussing the results from both the experiments reported in Papers III and IV and how the results from each experiment contributed to answering RQ2.1.

While Experiment 1 compared the differences in creativity levels between two groups who received desiderata framed as requirements and ideas, respectively, Experiment 2 compared creativity levels when desiderata were framed as prioritized requirements and ideas. Here, presenting desiderata as ideas meant framing desiderata as more ambiguous, unclear, and uncertain than presenting the same set of desiderata as more clear, rationalized, structured, and formal requirements specifications. This was done only to minimize the difference between the two groups as much as possible.

76
The results from both experiments suggest that framing desiderata as requirements or prioritized requirements significantly reduced design creativity (Findings C and D). The exploratory analysis suggests that fixation mediates the effect of problem structure on practicality (importance of specification is positively correlated with practicality: rho=0.297, p=0.005) but not on originality (importance of specification is negatively correlated with originality: rho= -0.144, p=0.108). This could either mean that fixation actually does not mediate the effect of problem structure on originality, or it could simply indicate that this effect is not very strong or the sample size of the experiment was not large enough to detect it. However, the literature suggests that fixation mediates the relationship between problem structure and originality and not practicality (see Chapter 2).

To summarize, both experiments indicate that presenting the same set of desiderata in slightly different ways has a profound effect on creative outcome. Presenting desiderata as formal requirements leads to less original but more practical designs, significantly undermining design creativity. It is, therefore, suspected that the high importance and high confidence connoted by the term requirement, augmented by a predetermined order of implementation—prioritized requirements—might impede designers’ ability to produce divergent ideas by promoting the view that the problem is clear, unambiguous, and already well understood. In short, requirements framing causes participants to fixate on the given desiderata instead of critically analysing the specifications for their content.

5.2.2 Answer to RQ2.2: What cognitive behaviours explain the negative relationship between specification formality and design creativity?

The qualitative data collected by executing a dialog-based protocol study were analysed in two separate phases, as explained in Section 3.2.4. The results from inductive process coding reveal seven different kinds of design actions performed by the participants while creating conceptual designs: uncritical acceptance, rejecting, considering and making design moves, grouping, questioning, assuming, and considering aesthetic qualities. By contrast, the results of the deductive closed coding analysis reveal a total of 1006 instances of fixation against 298 instances of critical assessment. Here, fixation refers to the design actions where participants 1) readily accept all the aspects of the task, such as requirements and priority levels; 2) adopt features and properties of known and existing examples without any reflection or evaluation; and 3) unreasonably reject
any divergent idea that deviates from the assigned task or existing examples. On the other hand, critical thinking consists of design actions whereby participants attempt to critically evaluate or reflect on the given task parameters or known examples. In a way, presenting desiderata as formal and structured requirements specifications hinders creativity by inducing fixation and restricting critical and divergent thinking.

While previous experimental research had demonstrated that presenting desiderata as formal requirements reduces creativity, the results of this exploratory analysis illuminate the underlying cognitive mechanisms that explain this relationship. This way, the answer to RQ2.2 is that fixation mediates the negative relationship between the formality of desiderata presentation and design creativity.

5.3 Summary of results

To summarize, the findings presented in this dissertation clearly show that presenting desiderata more formally as requirements affects design creativity. Specifically, when desiderata are framed in a more structured way as requirements or prioritized requirements, the result is less original but more practical designs, significantly undermining creative design performance.

These findings have several significant implications in the way desiderata are structured and presented in software engineering. The results directly conflict with the traditional and dominant view in RE that providing more clarity, precision, and structure to a problem is critical for software engineering success (Atkinson, 1999). Moreover, RE experts tend to reject the idea that more ambiguity and uncertainty would result in more creative solutions. In other words, when a project demands innovation—deviation from the usual approach—it might be beneficial not to over-structure or formalize desiderata as requirements. However, if a project requires more straightforward or archetypal solutions, presenting desiderata as requirements might be more appropriate than ambiguous specifications.

In RE literature, a requirement is something that is demanded, and it is often assumed that users already have a clear and precise set of requirements available to them. The only concern for requirements analysts is to elicit those requirements efficiently. However, in interdisciplinary user-centred design or human-computer interaction (HCI), a requirement is something that is needed, because software project success means that all (or at least most of) stakeholders should benefit from the system (Ralph & Kelly, 2014). Moreover, users do not answer questions (i.e., elicit requirements) by checking
an available directory. Instead, during an interview, the analyst and stakeholders co-construct an initial set of preferences (Lichtenstein & Slovic, 2006). Labelling a fleeting desideratum as a *requirement* can inflate both the confidence in it and its importance to a specific project. Giving such dubious requirements to a designer can eventually undermine project success.

Figure 9 presents a theoretical framework that explains the effect of specification formality on SE success—presenting desiderata more formally as requirements appears to induce fixation and hinder critical thinking, affecting software design creativity.

The main findings from all the studies reported in this dissertation, including the list of primary studies from Papers I and II, the originality and practicality grades from Papers III and IV, and the results from open coding analysis conducted in Paper V, are available online.

### 5.4 Threats to validity

The validity of empirical research in general may be threatened by several factors. Each publication in this dissertation includes a detailed discussion of the respective studies’ validity. The following section, however, summarizes the most critical validity threats according to the type of study by following the guidelines proposed by Wohlin et al. (2012).

#### 5.4.1 Systematic mapping study

Summarizing the knowledge concerning cognitive biases in SE research was not without some inherent limitations. First, the search string—`software' AND "cognitive bias"`,

---

4[https://figshare.com/articles/Results_Package/8342759](https://figshare.com/articles/Results_Package/8342759)
may have missed relevant articles that do not use the term “cognitive bias” in them (e.g., a study about confirmation bias that does not mention the term cognitive bias). Further, relevant articles concerning cognitive biases published in non-SE outlets may not have been indexed by the online search engines. Moreover, limiting our search to the computer science domain by using specific filters may have excluded relevant articles in other non-computer science domains (e.g., a study about cognitive bias in the SE context that was published in a psychology journal). To overcome these limitations, the following tactics were used to improve the external validity of this study:

- An iterative backward snowballing approach (Wohlin, 2014) was employed on the reference lists of the included studies.
- The profiles of the most active researchers were analysed to include any relevant studies that were otherwise missed.
- The first 150 results found by using the same search string on Google Scholar were analysed.
- Specific cognitive biases were searched based on specific terms (e.g., ‘software’ AND “emotional bias”) on all search engines and Google Scholar.

Despite these precautionary steps, some relevant studies might still have been missed. Notably, the unpublished or non-peer reviewed work, grey papers, and academic thesis and dissertations.

Nonetheless, synthesizing the research based on cognitive biases itself was problematic, due to two main reasons—1) The existing knowledge on cognitive biases is only some of the empirical generalizations concerning common SE errors and problems. Hence, the relationship between the biases, these errors, and their causes is unclear; and 2) the use of inconsistent terminologies to explain the biases, their causes, and their effects in primary studies hindered analysis.

Similarly, the search conducted to explore the state of creativity in SE research was intrinsically subjective, including the inclusion/exclusion process, quality assessment, and data extraction process. However, the analysis was conducted by the entire author team and followed established guidelines (Petersen et al., 2015). There also might be a threat to external validity, if a substantial number of studies related to innovation were unwittingly excluded to control the scope of the search.
5.4.2 Interviews

Threats to validity to the cross-sectional interview study were assessed for both construct validity and external validity.

Construct validity is concerned with the extent to which operational measures that are used in research represent what is investigated according to research questions (Runeson & Höst, 2009). In this thesis, construct validity deals with whether the interview questions were interpreted as intended and with the selection of interviewees. To address this issue, the entire author team developed the interview guide, following established guidelines (Whiting, 2008). Furthermore, before the actual interviews, three pilot interviews were conducted to validate the interview questionnaire and the guide.

External validity, meanwhile, is concerned with the extent to which the findings of the research can be generalized and the extent to which the findings are of interest to other cases (Wohlin et al., 2012). Since the interviewees were not randomly chosen, generalization of the results is not possible. However, a diverse sample, from varied roles and industries was sought to obtain a broader perspective. Finally, social desirability and the Hawthorne effect (Adair, 1984) may have biased interviewees’ responses, and the qualitative data were subjectively analysed although with complete agreement by the author team.

5.4.3 Experimentation

The findings from the two experiments reported in this dissertation must be considered in light of the limitations detailed below.

Regarding external validity, the participants were not chosen from a random population, so the statistical generalization of the results is not possible. Since the participants were not expert professionals but experienced post-graduate students, the results may not even generalize to novice or expert practitioners, although initial evidence shows that professionals and experienced students tend to perform similarly during experiments (Salman, Misirli, & Juristo, 2015). Moreover, as both experiments utilized a task comprised of a relatively disorganized list of desiderata, the results may not generalize to more refined requirements specifications. Additionally, running replications with more realistic or mission-critical design tasks might produce different results. The artificial laboratory setting may account for some confounding factors or unknown effects.
As far as construct validity is concerned, no objective scale to measure originality or practicality was found. Hence, subjective judgment by experts was used to assess both constructs; different judges might have produced different results. However, this threat was minimized by the establishment of standard assessment guidelines shared by both judges and piloting the evaluation procedure by selecting three random solutions from the set. Moreover, high inter-rater agreement in both experiments showed that the assessment procedure was highly reliable. Nevertheless, creativity is a multidimensional entity, and originality and practicality are not the only dimensions of creative design performance. Additionally, highly creative designs need not necessarily lead to better implementation and innovation. In terms of social threats, hypothesis guessing may have taken place because the participants may have become acquainted with the authors’ prior research. However, the hypotheses were never disclosed to the subjects before the experiments were conducted.

The controlled nature of the experiments suggests high internal validity. This was maintained by a detailed description of the experimental protocol and by the availability of the replication package (see Appendix 1), which contributes to robust reproducibility. The main objective of conducting the experiments was to establish a cause-and-effect relationship. The use of well-understood and strong statistical tests that meet their assumptions coupled with the analyses of the effect sizes contributed to high conclusion validity. The reliability of the treatment implementation should not be considered a threat per se, due to the design of the experiments. Mainly, the ideas framing (using the same set of desiderata) was used only to minimize the difference between the two groups. Additionally, the statistical tests’ assumptions were thoroughly assessed in each study, and no extraordinary events that could have influenced the course of the studies took place.

However, the findings from the meta-analysis are more prone to validity threats. Some of the data from the manipulation checks were non-normal with unequal variance. Multiple and imperfectly suited tests were used and therefore did not correct for multiple comparisons. Furthermore, combining data from two separate experiments can cause an unequal-groups threat to validity, which can be mitigated, or at least minimized, by running more replications. For all these reasons, the manipulation check may not correctly reflect the presence of fixation, which was further explored by conducting a protocol study. The findings are therefore presented as an exploratory analysis to differentiate them from the main findings of the two experiments.
5.4.4 Protocol study

Although protocol analysis is currently the most efficient method to explore and analyse the cognitive processes of the designers under investigation (Hughes & Parkes, 2003), it had the following limitations:

- When asked to speak aloud, participants could have altered their cognition and actions in ways which are neither predictable nor detectable. This threat was mitigated by running the protocol study with pairs of participants instead of individuals.
- A protocol study is a non-statistical research method that makes use of a particular task in a particular environment. Moreover, the participants were not chosen randomly. Therefore, the findings cannot be statistically generalized to other tasks, environments, or even populations, including practitioners.
- A protocol study involves making subjective inferences from the verbalized cognitive process of the participants, posing a serious threat to conclusion validity. Such ambiguity can be minimized by using brain scanning technologies to cross-check these inferences. However, no such approach was used to maintain the scope of the study.
- Analysing qualitative data is itself inherently subjective. This threat was mitigated by having the entire team review the coding technique, which resulted in multiple revisions, clarifications, and corrections.
- Some participants often conversed in a language other than English or were reluctant to think aloud. This threat was minimized by the facilitator’s prompting the participants periodically, if and when required. However, this prompting and their inability to converse in English may still have affected the data.
- Finally, while refining the conceptualization of the themes, the author team often renamed themes, merged multiple themes, or simply dropped unwanted themes and their corresponding codes. Despite much modifications, the themes and the relationship between the codes stabilized quickly and remained stable.
6 Conclusion and future research

This dissertation contributes to scholarship by highlighting the harmful effects of specification formality on software design creativity. The empirical data for this research were collected from controlled experiments and an exploratory protocol study, supported by secondary reviews and cross-sectional interviews. The main objective of this dissertation was to answer the research question: What is the effect of specification formality on design creativity? The synthesis of evidence suggests that specification formality triggers fixation, which impedes design creativity. While the results from controlled experiments showed that framing desiderata formally as requirements or prioritized requirements significantly reduces design creativity, the results of the protocol study showed that the negative relationship between specification formality and design creativity is mediated by fixation.

Contributing to the previous research on design fixation (Jansson & Smith, 1991), this dissertation proposes the concept of requirements fixation, defined as “the tendency to disproportionately focus on desiderata that are explicitly framed as formal requirements” (Mohanani, Turhan, & Ralph, 2019, p. 9). While previous research on design fixation shows that designers tend to fixate on features of the given examples, the results of this dissertation suggest that designers may also fixate on the way in which a problem is presented or communicated. Specifically, more formal and structured framing of desiderata can trigger requirements fixation that leads designers to attribute undue confidence and importance to desiderata presented as formal requirements statements, thus inhibiting their creativity. In other words, designers tend to not critically evaluate structure of a given task when that structure is presented as formal requirements.

To summarize, the findings of this dissertation highlight the fundamental conflict between creativity and requirement specifications. They show that a specification is presented may be as critical as its content. Satisfying all the requirements might not be, after all, the primary dimension of software engineering success.

Figure 10 and Table 10 summarize the research gaps, research questions, publications, main findings, and contributions reported in this dissertation.
6.1 Summary of contributions

This section discusses the four main contributions of this dissertation.

a) A constant and sustained interest in bias-related research is clear, with a scarcity of empirical research on debiasing techniques and a lack of established theoretical frameworks to explain and interpret the cognitive biases (Paper I).

The main objective of performing an SMS was to summarize existing SE research concerning cognitive biases to identify which cognitive biases are investigated in what SE activities by focusing on the bias antecedents, effects, and debiasing techniques. This secondary review supported the primary objective of this dissertation by identifying the relevant SE research literature to position future empirical investigations concerning fixation and framing effect. Another aim of conducting a systematic review was to aggregate the available knowledge on cognitive biases and reveal patterns that could explain the relationship between fixation and framing effect and whether any other cognitive biases affect creativity due to specification formality.
Table 10. Summary of the dissertation’s research gaps, research questions, and papers.

<table>
<thead>
<tr>
<th>Type</th>
<th>ID</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Gap</td>
<td>1</td>
<td>Lack of a comprehensive body of knowledge that summarizes the available literature in SE research involving cognitive biases, particularly concerning fixation and the framing effect.</td>
</tr>
<tr>
<td>Research Gap</td>
<td>2</td>
<td>Lack of a shared understanding between SE research and practice concerning the way creativity is conceptualized and assessed in SE.</td>
</tr>
<tr>
<td>Research Gap</td>
<td>3</td>
<td>The question of whether or not more structured desiderata framing (i.e., specification formalism) impedes design creativity.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>RQ1</th>
<th>What is the current state of research concerning cognitive biases and creativity in SE?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RQ2</td>
<td>How does specification formalism affect design creativity?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Publication</th>
<th>I</th>
<th>Cognitive Biases in Software Engineering: A Systematic Mapping Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>II</td>
<td>Perceptions of Creativity in Software Engineering Research and Practice</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>Requirements Fixation</td>
</tr>
<tr>
<td></td>
<td>IV</td>
<td>Requirements Framing Affects Design Creativity</td>
</tr>
<tr>
<td></td>
<td>V</td>
<td>Requirements Fixation: A Dialog-Based Protocol Study</td>
</tr>
</tbody>
</table>

| Main Finding      | A   | The study identified 65 primary studies investigating 37 cognitive biases. Most research is focused on investigating the antecedents and effects of biases. While the most frequently investigated cognitive biases belong to interest and stability bias, social and decision biases are least investigated. Significant confusion exists concerning the interpretation and theoretical conceptualization of these biases. |
|                   | B   | SE researchers and practitioners appear to perceive creativity differently, with minimal consensus, which hinders the development of evidence-based techniques and approaches to measure and enhance creativity. |
|                   | C   | Framing desiderata as requirements significantly reduced design creativity.                                                                |
|                   | D   | Framing desiderata as prioritized requirements significantly reduced design creativity.                                                     |
|                   | E   | Participants uncritically accepted all the aspects of the task (e.g., requirements, priority levels), adopted features of known examples without any critical reflection, and reject divergent ideas. |

| Contribution      | a   | A constant and sustained interest in bias-related research is clear, with a scarcity of empirical research on debiasing techniques and a lack of established theoretical frameworks to explain and interpret the cognitive biases. |
|                   | b   | Creativity is perceived as a cognitive process that produces original and practical ideas, and a creative artefact can be assessed subjectively by multiple experts in terms of the perceived novelty or quality of the idea. |
|                   | c   | The evidence suggests a negative relationship between specification formality and design creativity.                                         |
|                   | d   | The negative relationship between specification formality and design creativity is mediated by fixation.                                    |
The findings reported in Paper I showed that SE research often investigates the causes and effects of cognitive biases but rarely investigates techniques preventing biases or mitigation their effects. Moreover, none of the empirical studies empirically evaluated the existing debiasing approaches in the context of specific SE problems. Secondly, the psychological and sociological mechanisms underlying many cognitive biases are poorly understood. Most studies investigating the same cognitive biases are found to be disconnected from one another, with no reference to other similar studies. Better integration of studies and adapting (or developing) theoretical frameworks to explain the manifestation of biases would help create effective debiasing approaches. Finally, controlled experiments were the most preferred research method to investigate cognitive biases. Conducting more qualitative and exploratory research is necessary to understand how and why cognitive biases affect SE practices and how practitioners could perceive debiasing techniques. This study contributed to scholarship by providing a motivation to better understand the theoretical foundations underlying common SE problems and errors and develop interventions that could mitigate the biases and their corresponding issues.

b) Creativity is perceived as a cognitive process that produces original and practical ideas, and a creative artefact can be assessed subjectively by multiple experts in terms of the perceived novelty or quality of the idea (Paper II).

To assess creativity, it was important first to understand how creativity is perceived in SE. A multi-methodological study comprised of an SMS with cross-sectional interviews was conducted to check whether there exists any shared consensus between SE research and practice about the way in which creativity is understood, assessed, and improved. The study found that SE research and practice adopt very limited views of creativity—limited to a product’s creativity—while ignoring other perspectives of place, process, person, potential, and persuasion. Moreover, it illuminates a research-practice gap where practitioners are unaware of the state-of-the-art techniques available to assess and enhance creative performance. This study makes two main contributions to this dissertation:

– A consensus that creativity is a cognitive process that produces not only original but also practical ideas
the most popular creativity assessment technique is a subjective assessment of creative artefacts (products) by multiple experts.

c) The evidence suggests a negative relationship between specification formality and design creativity (Papers III and IV).

The results from Papers III and IV indicate that presenting the same set of desiderata in a slightly different manner has an overwhelming effect on creative design performance. Specifically, presenting desiderata as requirements or prioritized requirements significantly reduces design creativity. This experimental research contributes to the dissertation as follows:

- Framing desiderata as more structured and formal requirements results in less original but more practical designs, significantly undermining design creativity.
- Designers are highly affected and susceptible to minor changes in the vernacular used to communicate the desiderata; by simply using the terms like requirements and ideas can deleteriously affect designers’ creative potential.
- Software practitioners should recognize, analyse, and accept the ambiguity, uncertainty, and conflicts in system desiderata as unavoidable and potentially beneficial for creative solutions.

To summarize, this part of the research highlights the potential power of minor changes in the way in which desiderata are communicated and the sensitivity of designers to cognitive biases like fixation and framing effects.

d) The negative relationship between specification formality and design creativity is mediated by fixation (Paper V).

Exploring the cognitive processes that cause desiderata framing to affect design creativity was essential to explaining and understanding the existence and manifestation of fixation. Combining randomized controlled trials with an exploratory study could have helped to better understand the mechanisms that reduce design creativity. The results from Paper V show that designers tend to execute design actions associated with fixation more frequently than design actions associated with critical thinking. Specifically, designers do not critically evaluate and readily accept the a given task structure when it is presented more formally. This paper contributes to the dissertation in the following ways:
– It provides an initial taxonomy of the software design actions executed by designers while creating software designs.
– It shows that the negative relationship between specification formality and design creativity is mediated by fixation.
– It proposes a theoretical framework (see Figure 9) explaining the relationship between specification formality and software design creativity.

To summarize, the observations from the exploratory research suggest that framing desiderata more formally as requirements specifications triggers requirements fixation, which impedes the critical thinking ability of designers.

6.1.1 Implications for SE education

The notion that managers and analysts elicit system requirements and that designers and developers convert those requirements into a system design is simply misleading. Overusing the term requirements for desiderata that are less important and uncertain results in over-simplification and over-rationalization of a complex phenomenon that undermines the diversity of the potential desiderata. In this way, SE education continues to portray over-rationalized and over-simplified views of RE. Traditional RE practices obscure the conflicts, disagreements, ambiguity, and diversity in software projects. Such practices lead to an inaccurate understanding of how desiderata are perceived, which impedes design creativity. Moreover, the ACM/IEEE official curriculum for undergraduate programs in software engineering barely mentions the importance of creativity in SE activities and conceptual design generation (e.g., Ralph, 2012). This dissertation suggests several improvements to SE education:

– The outdated RE curriculum should be updated to include other important topics like human-centred design, ways to write user stories, and maintaining a product backlog (e.g., Sedano, Ralph, & Péraire, 2019).
– Non-empirical theories and concepts like the Waterfall Model and Project Triangle should be replaced by concepts with a stronger evidentiary basis (Ralph, 2018).
– Students should be given more open-ended assignments with unclear goals, conflicting stakeholder opinions and choices, ill-structured problems, and incomplete specifications. This would prepare them for more realistic projects and problems. Moreover, students should be exposed to theories and concepts, such as actor-network theory (Law & Hassard, 1999) and soft systems methodology (Checkland & Poulter,
2006), that would help them understand ambiguous contexts, conflicting stakeholders, and fewer formal specifications.

- SE curricula could benefit from teaching many underrepresented topics such as design thinking and theories of cognitive biases. However, it may be more efficient to teach the concept of cognitive biases in the context of common SE problems rather than as a stand-alone topic.
- More extensive training in creativity-enhancing techniques and approaches would help students to come up with alternative solutions to a problem.

To summarize, SE education should avoid an over-simplified and over-rationalized version of software engineering. Students should be exposed to more realistically imperfect requirements, different dimensions of software engineering success (Ralph & Kelly, 2014), taught creativity-enhancing techniques, and well trained to distinguish real requirements from the dubious ones.

### 6.1.2 Implications for SE practice

For SE practitioners, the empirical investigations reported in this dissertation primarily suggest that over-using the term requirement for all desiderata can impede creativity, irrespective of the contents of the specifications themselves. Software engineering success is a multidimensional phenomenon comprising of many critical aspects such as creativity, artefact quality, market performance, efficiency and impacts on stakeholders over time (Ralph & Kelly, 2014). A more formal list of specifications could simultaneously result in a less creative solution and higher quality code. In a way, software engineering success is a complex construct where practitioners and managers should adopt a comprehensive view of software engineering, consider trade-offs between various dimensions of success, and avoid any misleading and dubious methods or simplified specifications. However, desiderata should be made less formal and framed to induce scepticism, if more innovative designs are preferred. In contrast, a clear, formal, and a well-structured list of requirements might help in high correctness of code, which can increase the possibility of success of a software developed for mission- or safety-critical domains.

Specifically, the results of this dissertation show that the high confidence and importance connoted by the term requirement shuts down designers’ creative abilities and potential by encouraging the view that the problem is well understood, well
structured, and—indeed—already solved. Therefore, if more creative solutions are preferred, desiderata must be presented less formally to promote a healthy scepticism.

Moreover, if stakeholders disagree on a problem that the system should solve, designers and developers should attempt to model these disagreements and conflicts instead of documenting an over-rationalized version of the problem. Similarly, managers and analysts should avoid labelling, and accepting, all desiderata as requirements, which might overstate the confidence and importance of a specific desideratum. Although the research in this dissertation used a list of ideas framing to compare creativity, requirements specifications must not be renamed ideas specification. Instead, practitioners must speculate which way of representing desiderata (e.g., use-cases, user stories, personas, scenarios) is most appropriate for the project under consideration.

Furthermore, the findings might even encourage practitioners to consider two properties of each desideratum—confidence (Nolan, Abrahão, Clements, & Pickard, 2011) and importance (Karlsson, 1996). While confidence refers to the certainty of the desideratum’s relevance, importance refers to how crucial a particular desideratum is for success. Framing low-importance and low-confidence desiderata as requirements is more tricky than labelling high-importance, high-confidence desiderata as requirements. Hence, when a desideratum is not very important or is uncertain, practitioners should consider it as a “preferred future feature” or even as an “unvalidated idea”, rather than “optional requirements” or “low-priority requirements”. However, presenting desiderata more formally might result in more practical designs. More generally, to encourage creative and innovative solutions, the term “requirements” must be reserved for desiderata have high importance and with enough evidence to support the notion of requirement.

Finally, the findings from the secondary review of cognitive biases suggest that adopting debiasing practices can help mitigate many cognitive biases quickly and with low overhead costs. Debiasing tasks is often easier than debiasing people (Fischoff, 1981). For instance, anchoring and adjustment bias may affect a team estimating user-stories, but using the Planning Poker technique improves the estimation ability of the team (Haugen, 2006) as an alternative to training the entire team to understand and resist anchoring and adjustment bias. Creativity can be fostered by running periodic creativity-enhancing workshops (Maiden, Gizikis, & Robertson, 2004), during which designers and developers could be explicitly asked to produce out-of-the-box ideas. The entire software development team can be run through multiple lateral-thinking exercises to improve their creative potential (De Bono, 2010).
6.2 Future research avenues

From a research perspective, the results of this dissertation open multiple opportunities for future studies.

More research is required to explain the relationship among fixation, desiderata framing, and creativity. Several controlled experiments and replications with novice and expert designers as participants, combined with confirmatory field studies, are needed to investigate further how desiderata framing affects creative design performance. Similarly, a family of experiments involving more complex and mission-critical software projects would help provide further insights into the relationship between specification formality and creativity. While traditional RE research focuses more on prioritizing desiderata and distinguishing between mandatory and optional or between functional and non-functional requirements, future RE research may also benefit from developing techniques to indicate the epistemic status of a desideratum (e.g., 80% certainty that the system under consideration will need to support a secure payment gateway). The results of both sets of experiments suggest presenting less important and less certain tasks in a way that promotes scepticism but is still appropriate for a particular project. On the other hand, it would be interesting to investigate the effects of giving a desideratum low confidence and low importance and still labelling it as a requirement.

RE research over the years has proposed multiple ways to represent desiderata, such as user stories (Schwaber, 2004), use cases (Cockburn, 2000), and requirements specifications (IEEE, 1998). However, for any evidence-based recommendations concerning which representation to use, techniques should be compared directly using controlled experiments. RE practices can also benefit from developing tools, techniques, and approaches for representing, analysing, and addressing ill structured design tasks and managing ambiguity and conflicts between stakeholders. Moreover, future SE research should develop ways to present desiderata in such a way that designers envisage solutions that are highly creative: designs that are simultaneously original and practical, not one or the other.

From the point of view of practitioners, however, the ability to produce creative software depends on the way they critically evaluate a problem situation. SE research could thus investigate how combining importance and confidence metadata could affect originality and practicality of designs. In the context of this dissertation, the previous literature indicated that the formal presentation of desiderata triggers fixation. However, there might be other cognitive biases in play or other cognitive mechanisms...
that might be associated with fixation: conceptual fixation (Youmans & Arciszewski, 2014), confirmation bias (Calikli & Bener, 2010), selective attention (Lavie, 2005), and miserly information processing (Stanovich, 2009). Future research in SE must look forward to exploring these and other similar phenomena that affect practitioners’ judgment and creative abilities.

In a broader perspective, the widespread confusion encountered while analysing the current state of bias-related literature warrants multiple lines of investigations: 1) more qualitative studies, including protocol analysis (Ericsson & Simon, 1993), grounded theory (Charmaz & Belgrave, 2007), and case studies (Yin, 2017) must be conducted to understand the complex social and psychological mechanisms from which cognitive biases emerge; 2) a better understanding of these mechanisms would inform the development of new debiasing techniques, while the existing ones can be evaluated using controlled experiments (Wohlin et al., 2012) or action research (Sein, Henfridsson, Purao, Rossi, & Lindgren, 2011) that suits specific software contexts (e.g., evaluating Planning Poker (Moløkken-Østvold, Haugen, & Benestad, 2008) to reduce errors during cost and effort estimations); and 3) SE research requires a better taxonomy to categorize and classify biases according to their causal mechanisms, the nature of their effects, or the way in which they manifest themselves in software projects. Such a classification schema could facilitate clearer reasoning about biases and possibly inform debiasing techniques as well.

Finally, as far as the role of creativity in SE is concerned, conducting more research by collaborating directly with industry is needed to develop objective assessment techniques to explore how creativity manifests in software development projects, how personality, creative potential, mental processes and the work (context) environment affect creative performance, and establishing the relationship between intrinsic and extrinsic motivation, autonomy, and creativity.
References


Bartelt, C., Vogel, M., & Warnecke, T. (2013). Collaborative creativity: From hand drawn sketches to formal domain specific models and back again. In *Moroco@ECSWC* (pp. 25–32).


for combinational creativity in requirements engineering. In 2014 ieee 22nd international requirements engineering conference (re) (pp. 243–252).


Cockburn, A. (2000). Writing effective use cases, the crystal collection for software professionals. *Addison-Wesley Professional Reading.*


Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A
typology and critical analysis of framing effects. *Organizational behavior and human decision processes*, 76(2), 149–188.


Mumford, M. D. (2003). Where have we been, where are we going? taking stock in creativity research. *Creativity research journal, 15*(2-3), 107–120.


Torrance, P. (1966). *Thinking creatively with pictures, booklets a and b; thinking creatively with words, booklets a and b*. Personnel Press.


Youmans, R. J. (2011). The effects of physical prototyping and group work on the


Appendix 1 Lab Package

A lab package for replicating the studies presented in this dissertation is available at:
https://figshare.com/articles/Replication_Package/8342807

This package includes:

- **Experimental Task Documents**: A description of the tasks used to conduct the studies in Papers III and IV, comprising of requirements specification document, prioritized requirements specification document, ideas specification document and copy of blank design template.

- **Protocol Study Task Document**: A description of the task used to conduct the protocol study in Paper V.

- **Questionnaires**: Pre- and post-study questionnaires administered to the participants before and after the experimental study. These documents were mainly used to record demographic information and the answers to the manipulation check questions.

- **Quantitative Data Analysis Script**: A description of the statistical tests used to analyse the data collected in Papers III and IV.

- **Interview Questions**: A description of the interview questions and the corresponding rationale used in Paper II.
Original publications


Reprinted with permission from The Institute of Electrical and Electronics Engineers, IEEE (I, II, and IV) and Association for Computing Machinery, ACM (III).

Original publications are not included in the electronic version of the dissertation.
721. Giunti, Guido (2018) 3MD for chronic conditions: a model for motivational mHealth design
727. Hurkainen, Sonja (2018) The roles of individual demographic history and environmental conditions in the performance and conservation of northern orchids
728. Haapalahti, Reijo (2019) Yksilön toimien vaikutukset aluekehitykseen: ammatilliseen perustutkintokoulutukseen liittyvät odotukset ja tulokset Pohjois-Pohjanmaalla
729. Tokkonen, Helena (2019) Say, Do, Make?: user involvement in information systems design
731. Mylonopoulou, Vasiliki (2019) MAD: designing social comparison features in health behaviour change technological interventions
732. Shevchuk, Nataliya (2019) Application of persuasive systems design for adopting green information systems and technologies
733. Tripathi, Nirnaya (2019) Initial minimum viable product development in software startups: A startup ecosystem perspective
Rahul Prem Mohanani

REQUIREMENTS FIXATION:
THE EFFECT OF
SPECIFICATION FORMALITY
ON DESIGN CREATIVITY