How Templated Requirements Specifications Inhibit Creativity in Software Engineering

Rahul Mohanani, Paul Ralph, Burak Turhan, Senior Member, IEEE and Vladimir Mandić, Member IEEE

Abstract—Desiderata is a general term for stakeholder needs, desires or preferences. Recent experiments demonstrate that presenting desiderata as templated requirements specifications leads to less creative solutions. However, these experiments do not establish how the presentation of desiderata affects design creativity. This study, therefore, aims to explore the cognitive mechanisms by which presenting desiderata as templated requirements specifications reduces creativity during software design. Forty-two software designers, organized into 21 pairs, participated in a dialog-based protocol study. Their interactions were transcribed and the transcripts were analyzed in two ways: (1) using inductive process coding and (2) using an a-priori coding scheme focusing on fixation and critical thinking. Process coding shows that participants exhibited seven categories of behavior: making design moves, uncritically accepting, rejecting, grouping, questioning, assuming and considering quality criteria. Closed coding shows that participants tend to accept given requirements and priority levels while rejecting newer, more innovative design ideas. Overall, the results suggest that designers fixate on desiderata presented as templated requirements specifications, hindering critical thinking. More precisely, requirements fixation mediates the negative relationship between specification formality and creativity.

Index Terms—Cognitive bias, software design, fixation, critical thinking, requirements, requirements engineering, protocol analysis.

1 INTRODUCTION

Different organizations initiate new software projects in many different ways. For teams developing consumer applications and enterprise systems, however, the initiation process often seems to include:

- speaking with prospective users and other stakeholders about their needs, wants, preferences, etc.;
- synthesizing stakeholders’ opinions into some documents;
- determining the main features that the product will have; and
- creating mock-ups and diagrams illustrating the main features and user interfaces.

These activities can be sequential, parallel, in different orders and performed by the same or different people. Nevertheless, it raises an obvious question—what kind of documents are best for synthesizing stakeholders’ opinions? Different areas of research seem to have reached different conclusions.

The more positivist side of requirements engineering (RE) research tends to assume that software projects have discoverable and documentable requirements, and that understanding these requirements is critical for designing good software systems [1], [2]. It seeks to elicit unambiguous, consistent, complete, feasible, traceable and verifiable requirements [3]. Good requirements specifications should lead to good software designs [4] because meeting requirements is what ‘good’ means.

Contrastingly, the more naturalistic side of RE, as well as research in human–computer interaction, user–centred design, and the interdisciplinary design literature tends to assume that:

- software projects do not have discoverable and documentable requirements (cf. [5]);
- stakeholders do not even have stable, retrievable preferences (cf. [6]); and
- products have numerous stakeholders who do not agree on the problem(s) to solve or how the product should solve them (cf. [7]).

Forcing vague, unstable, conflicting preferences into unambiguous, consistent requirements specifications encourages designers to converge prematurely on oversimplified problems and inappropriate solutions [8]. Eliciting templated requirements specifications should therefore lead to designs that satisfy contracts but not users, which is antithetical to user–centred design.

Our previous work showed that presenting a set of desiderata as templated requirements specifications led to less creative product designs than presenting exactly the same desiderata as uncertain ideas [9], [10]. However, experiments like these are not suitable to explore cognitive mechanisms underlying causal effects, so we have evidence that the presentation of desiderata affects creativity but we don’t know how. This raises the following research question.

Research question: How do fixation and critical thinking explain reduction in design creativity when desiderata are presented as templated requirements specifications?

Here, fixation means the tendency of the designers to pay excessive and undue attention to the given problem by...
readily converging on an available or known solution (cf. [11]); critical thinking is defined as “disciplined thinking that is clear, rational, open-minded, and informed by evidence” [12]. Design creativity, for our purposes, denotes the originality and practicality of new product concepts. Desiderata are properties of a real or imagined system that are wanted, needed or preferred by one or more project stakeholders [13]. We use this term because it helps us remember that the set of things a stakeholder wants and the set of things needed for a system to succeed do not always coincide. While requirements specification is defined as “a statement that identifies a capability or function that is needed by a system in order to satisfy its customer’s needs” [14], templated requirements specification (TRS) is requirements specification written in a specific syntactic structure using a restricted (controlled) natural language [15], in this case, for example, “The system shall facilitate diet planning”.

Next, we review existing literature (Section 2). Then, we describe our research design including data collection and analysis (Section 3), followed by the results (Section 4). Section 5 discusses the theoretical framework and summarizes the study’s implications and limitations. Section 6 concludes the paper with a summary of its contributions.

2 Background

This section summarizes the major concepts involved in this study: desiderata, task structuring, creativity and fixation.

2.1 The concept of desiderata

In the most extreme positivist view of RE, requirements are a property of the environment, which are motivated by the desires, wants and needs of the stakeholders [16, 17]. Requirements analysts elicit them, and success means fulfilling them. Since social reality is real and objective, requirements exist in the world, waiting to be discovered. Project success means meeting and satisfying the requirements.

In the naturalistic view of RE and the constructivist view of product design, requirements do not exist in an objective reality [5, 18, 19] waiting to be discovered. Reality is socially constructed. Stakeholders usually have unstable, unreliable, conflicting desiderata [6, 20] and use different processes and representations to express their opinions. Research in RE has tried to organize and manage these conflicts and inconsistencies among stakeholders to elicit requirements. Some of the techniques include the ViewPoints—a framework that facilitates capturing, representing and organizing multiple stakeholders’ viewpoints and perspectives [21], conflicts and goal-modelling [22] and AbstFinder—a tool that helps finding abstractions and textual ambiguity in natural language text [23]. Project success then means delivering benefits to stakeholders [24]. Fundamentally, RE is about establishing a balance between the positivist approach where desiderata are considered as a singular truth embodied in the form of a formal specification, and the naturalistic approach where requirements are a product of the conflicts and contradictory viewpoints of the stakeholders involved (cf. [25]).

Practically, the main difference in the above discussion is what happens when a stakeholder demands that the system has some property or fills some need. In the positivist view, the optative speech act of demanding manifests a requirement [26]. In the multi-perspective constructivist view, the stakeholder’s demand is just an opinion of note; the demanded property may be a necessary condition for success, irrelevant, or even prevent success. This manuscript assumes the latter.

2.2 Task structuring

Problems are often conceptualized on a spectrum of well-structured 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” [27, p. 65]. A body of empirical research shows that task structure is negatively associated with design performance (cf. [28] for summary). “Over-concentration (over-structuring) on problem definition does not necessarily lead to successful design outcomes” [29, p. 439] for at least four reasons:

1) less specific goals reduce cognitive load [30], which leads to more learning;
2) less specific task framing results in more creative solutions [31];
3) designers often fixate on experience [32] or on an initial set of ideas [33]; and
4) designers often process whatever little information they have and quickly assimilate it into the problem schema, improving their understanding of the problem [34].

Perhaps unsurprisingly, then, our recent experiments showed that presenting desiderata as TRS reduced creativity [9, 10]. We hypothesized that presenting desiderata as TRS triggers a specific cognitive bias, which we call requirements fixation. Fixation broadly refers to the tendency to “disproportionately focus on one aspect of an event, object, or situation, especially due to self-imposed or imaginary obstacles” [11, p. 5]. Requirements fixation, then, is the tendency to attribute undue confidence and importance to desiderata presented as TRS. We use the term requirements fixation to allude to similarity to design fixation: the well-established tendency for designers to generate solutions very similar to given examples [35] or existing artifacts [36].

Although the precise mechanism by which increasing task structure reduces design performance and creativity remains unclear, design expertise seem to moderate the relationship. Expert designers tend to resist initial problem framing (e.g. given TRS) and proceed via an improvised, solution-focused approach [37]. Expert designers consider all problems ambiguous and ill-structured, focusing on solution generation rather than analyzing the given problem [29, 38]. On the other hand, novice designers often treat ill-structured problems as well-structured, thereby compromising the potential for creative solutions [39]. However, recent research suggests that novice designers as compared to
more experienced designers, are less fixated on the problem
domain and, hence, are able to generate highly creative solutions [40], [41].

2.3 Creativity
RE is a creative process [42], [43], where analysts and multi-
ple stakeholders collaborate to make sense of a problematic
situation and conceptualize a common mental model of a
possible system [44]. In RE, creativity enhancing workshops (e.g., [45]) are extensively used to provide clarity
for requirements identification [46], [47] and generate novel
and creative requirements [48], [49]. In these workshops,
creativity is often linked with divergent thinking [50], i.e.,
exploring multiple and diverse solutions to a given problem.
While Brainstorming helps in generating most number of
requirements, Hall of Fame approach (viz., [51]) helps in
developing multiple creative requirements [52]. In another
study, interactive collaboration techniques employed during
the workshops helped in generating more creative require-
ments [48]. In a nutshell, requirements can be understood
as entities that encapsulate the results of creative thinking
about the system being developed [53].

Research in RE also focuses on ways to discover and
generate creative requirements by leveraging the way in
which problem situations and early ideas are represented.
One study preferred using user stories to explore novel
ideas, which were then used to measure the personality
traits and creative potential of the participants influencing
their creative abilities [54]. In another study, high-level
goals, represented as goal models (viz., [55]), were combined
with creativity enhancing techniques to explore and gen-
erate creative requirements [56]. Integrating the concept
of combinational creativity (viz., [57]) with use cases (e.g., [58])
and creativity enhancing framework (e.g., [59]) are also used
to discover and develop creative requirements.

Creativity, itself, is a poorly-understood, multi-
dimensional construct [60], [61]. However, creativity research
may be organized into the “Six P’s” (viz., [62], [63]), as
follows: 1) creativity’s underlying cognitive process; 2) the
creative product; 3) the person (or personality) doing the
creative work; 4) the place (or context) of the creative work;
5) stimulating creative thinking (or persuasion) and 6) im-
proving creative potential.

Here, we are primarily concerned with product creativity.
Product creativity entails two dimensions: (1) novelty or
originality [64], [65] and (2) practicality or usefulness [66],
[67]. Therefore, we conceptualize creativity as the ability
of a designer to choose both novel and practically useful
features, graphical elements and aesthetic properties of a
software system.

2.4 Fixation
Cognitive biases are systematic deviations from optimal rea-
soning [68]. Since software engineering involves lots of rea-
soning, cognitive biases help to explain common problems
in software design [69], [70], testing [71], [72], requirements
engineering [73], [74] and project management [75].

Fixation is a cognitive bias in which “blind adherence
to a set of ideas or concepts limit[s] the output of concep-
tual designs” [32]. Several experiments have demonstrated
design fixation—the tendency for designers to generate sol-
lutions very similar to given examples [32], [33] or existing
artifacts [36]. The cognitive mechanisms underlying design fixation are not well understood. However, [35] suggest that
designers may fixate on a known but limited set of ideas or
an existing body of knowledge, and classify design fixation
into three broad categories:

1) Unconscious adherence—designers depend too
heavily on ideas encoded in long term memory,
sometimes due to heavy load on working-memory.

2) Conscious blocking—designers dismiss new ideas
due to over-dependence and confidence on their old
ideas or past experience.

3) Intentional resistance—designers intentionally res-
ist new ideas to designs that were previously suc-
cessful.

Meanwhile, design fixation is moderated by several factors:

1) The domain; for instance, mechanical engineers fix-
ate more than industrial designers [76].

2) examples—common examples causes more fixation
than unusual or rare examples [77];

3) defixating instructions—explicit instructions to
avoid features of given or existing artifacts [78];

4) providing good quality examples than flawed or no
examples lead to better design performance [79];

5) product dissection activities [80]; and

6) physical prototyping [81].

SE research shows that inconsistency in requirements
specifications may reduce premature commitment [82].
More generally, the way a task is communicated or pre-

tented may also cause fixation [83], [84]. This is related
to the framing effect—“the tendency to give different re-
sponses to problems that have surface dissimilarities but are
formally identical” [85, p. 88]. Desiderata can be presented
in many different forms, including TRS, personas, scenarios,
use cases and requirements statements. We can think about
these forms as different ways of framing a design task, and
task framing affects design performance [3], [10], [86].

While, the underlying cognitive mechanisms that re-
duces creativity of design concepts due to TRS are not yet
well explored, our recent experiments in SE (e.g., [9], [10])
suggest a typical behavior where designers tend to shut
down their creative potential by further inflating the high
importance and confidence connoted by the TRS.

3 RESEARCH DESIGN
3.1 Dialog-based protocol analysis
To answer our research question, we need real-time insight
into software designers’ cognitive processes. Think-aloud
protocol analysis is a research methodology in which partici-
pants verbalize thought sequences. Researchers analyze par-
ticipants’ words for insight into their thinking. Researchers
assume that “any concurrent verbalization produced by a
subject while solving a problem is a direct representation
of the cognitive functioning (i.e., mental processes) of the
subject’s working memory” [67].

However, verbalizing our thought process probably
changes those thought processes in imperceptible ways.
We can mitigate this limitation using dialog-based protocol analysis [88], in which participants work in pairs or groups, and we analyze their natural dialog instead of a forced monologue.

Protocol analysis is common in design studies [89], psychology (e.g. [87]), medicine (e.g. [90]) and software engineering [88], [91]. We see protocol studies as most consistent with a critical realist philosophy of science (cf. [92]).

3.2 Purpose and scenario

The purpose of this study is to determine why presenting desiderata as TRS reduces design creativity. To do so, we observe participants in a simulation of a situation where a software team is given a set of TRS and then asked to design an appropriate application. Although RE and high-level designing are increasingly merged (e.g., [93], [94]), such situations arise often in outsourcing arrangements (where the outsourcer provides a questionable requirements specification and expects the team to develop mock-ups without much access to prospective users or client representatives). In our experience, many software teams that have stakeholder access still have specifications forced upon them by clients, management, marketing or other stakeholders. Fig. 1 summarizes our research protocol.

An alternative question—whether modeling desiderata as TRS harms creativity in teams that both create the TRS and then design the product—is potentially fruitful avenue of future work. We did not attempt such a simulation because it is inconsistent with the prior experiments we are attempting to re-examine and entails numerous unresolved methodological problems [96] section 1.3.

3.3 Participants and pairing

We recruited a convenience sample of 18 professional software developers (14 men, 4 women) from Company X, which develops web and mobile applications for government agencies, corporations and educational institutions. Company X has 30 employees, with a typical project duration of 2-3 years. We selected Company X because it was willing to participate due to close ties to one of the authors. These participants had a mean age of 31 years (σ = 5.65).

Meanwhile, we recruited 24 post-graduate students (21 male, 3 female) enrolled in the information processing science program at the first author’s university. Student participants had a mean age of 23 years (σ = 6.07). They received extra credit in one of their courses for participating.

While professional participants had a mean work experience of 2.3 years. All participants had at least 1 year of experience in software design. However, none of the participants had any experience with developing health and fitness applications—the domain used here. Participants were paired based on availability. Each pair comprised either two professionals or two students.

3.4 Execution of the study

The study was approved by the University of Auckland Human Participants Ethics Committee (UAHPEC), New Zealand. We facilitated and supervised the data collection from September to December, 2016, at Company X’s office for the professionals and at the students’ university. Every participant-pair was scheduled an individual session in a quiet room. On arrival, the study was described and participants signed a consent form and completed a demographic questionnaire.

We used the same task as our previous experiments [9], [10]. The task document listed 25 desiderata presented as TRS and organized into five priority levels: high, high–medium, medium, medium–low and low. Each TRS began with “The system shall” (consistent with [3]) and phrased as, for example, “The system shall measure calorie intake”, “The system shall recommend activities” and so on. Crucially, many of the desiderata presented in this task are ill-considered, inconsistent or over-complicated. The TRS was compiled to engender skepticism.

The participants were also given identical design templates comprising blank, mobile screen-sized boxes in portrait and landscape orientations with space for written explanations. Participants were then asked to generate conceptual designs of a health and fitness mobile application. Participants could use as many templates as needed. The participants were encouraged to discuss their thoughts while creating designs.

The first author acted as facilitator for both groups. Sessions were limited to 60 minutes at Company X’s request. Students were also limited to 60 minutes for consistency. During each session, the facilitator took notes, reminded participants to stick to English and prompted participants who made design moves without discussion.

We piloted the study once to check the recording equipment. No subsequent changes were made to the task or the study procedure. All task documents including the TRS and their qualities are available in our replication package (see Section 7).
We envisaged the data analysis in two phases: 1) inductive process coding to explore the cognitive mechanisms used by the participants; 2) deductive closed coding using concepts from existing literature to re-analyze the data through a specific theoretical lens. We used NVIVO (www.qsrinternational.com) to organize, analyze and visualize the qualitative data. The coding process is briefly described as follows:

### 3.5 Data collection and analysis

The sessions were transcribed by the first author. Although we did not correct participants’ grammar or malapropisms, we removed verbal static (e.g., “um”, “ah”, “uh”). We refer to each transcript by a unique identifier starting with a ‘P’. P1–P9 are the professionals; P10–P21 are the students.

We envisaged the data analysis in two phases: 1) inductive process coding to explore the cognitive mechanisms used by the participants; 2) deductive closed coding using concepts from existing literature to re-analyze the data through a specific theoretical lens. We used NVIVO (www.qsrinternational.com) to organize, analyze and visualize the qualitative data. The coding process is briefly described as follows:

#### 1) The first author performed process coding of all the transcripts (see Section 3.5.1).

#### 2) The second and the third author audited process coding.

#### 3) The auditing resulted in renaming some of the codes (e.g. doing design was renamed to making moves); some codes were combined together or rejected altogether (e.g. avoiding options and rejecting moves were combined to rejecting design moves).

#### 4) The first and the second author performed closed coding of one transcript together to ascertain the coding scheme (see Section 3.5.2).

#### 5) The first author then coded the rest of the transcripts.

<table>
<thead>
<tr>
<th>Theme</th>
<th>Example labels</th>
<th>Example quotations and dialogues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Making design moves</td>
<td>Discerning design moves</td>
<td>“OK, you’re on the intake page or seeing just the intake, and then you will have a small button.” “A-ha, so it could be like this...” “Icon, something like this...” “Yeah” (P15)</td>
</tr>
<tr>
<td></td>
<td>Generating multiple design options</td>
<td>“We can add into the ‘diet screen’ an option to recommend.” “Or rather it should be over here...you select and you search.” (P2)</td>
</tr>
<tr>
<td></td>
<td>Making moves</td>
<td>“We can have help me button to propose exercise. And eat this meal or skip, best practice to reach goal” “The BMR is calculated after this segment screen you have all the information” (P2)</td>
</tr>
<tr>
<td>Uncritically accepting</td>
<td>Accepting features of existing examples</td>
<td>“You have all the history, in one click. And here you have, most of the fitness apps like FitBit have this, so, why not use it?” “Yeah, okay” (P2)</td>
</tr>
<tr>
<td></td>
<td>Accepting priorities</td>
<td>“If we have to design an app that shows all these five requirements that are on a high priority, we would probably want to have all the data on one screen” (P16)</td>
</tr>
<tr>
<td></td>
<td>Accepting requirements</td>
<td>“So now we have requirement number 5, to provide workout history and performance analysis” “Yeah, that could be on the main page that given history and performance analysis” (P11)</td>
</tr>
<tr>
<td>Rejecting</td>
<td>Rejecting design moves</td>
<td>“I just thought we could split the screen and it shows you the current activity and any previous activities” “No, it can still show the previous one. Let’s stick to this way for now” (P16)</td>
</tr>
<tr>
<td></td>
<td>Rejecting requirements due to no knowledge</td>
<td>“So on to requirement 10, we shouldn’t consider much on this one?” “Yeah... it’s more like a system requirement, I don’t know...let’s move on” (P13)</td>
</tr>
<tr>
<td></td>
<td>Rejecting specifications due to time constraints</td>
<td>“Well the analysis screen would be the most complicated one. I don’t think we have time for that now” “Yeah, let’s check other ones” (P6)</td>
</tr>
<tr>
<td>Grouping</td>
<td>Grouping seemingly similar requirements</td>
<td>“So recommend recipes and recommend workouts and recommend diet food, all could be grouped as one. They all say recommend” “Yeah, okay” (P11)</td>
</tr>
<tr>
<td></td>
<td>Grouping as input-type data</td>
<td>“We put a dashboard feature” “like for input data” “Yeah, let’s group them as input data” (P1)</td>
</tr>
<tr>
<td></td>
<td>Grouping requirements of same priority level</td>
<td>“The first two high priority tasks of measuring calorie intake and what user eats and drinks is the same thing” “That’s true” “So those are kind of easy to put together...” (P15)</td>
</tr>
<tr>
<td>Questioning</td>
<td>Questioning priority levels</td>
<td>“The system must allow the user to plan workout...this isn’t very important. Shall track speed and distance...this is very important” “Me too” (P21)</td>
</tr>
<tr>
<td></td>
<td>Questioning examples</td>
<td>“They FitBit app have sensors and they can do precise measuring...running is based on GPS.” “Yeah, maybe we should do something else. GPS is not always good and reliable.” (P3)</td>
</tr>
<tr>
<td></td>
<td>Questioning requirements</td>
<td>“We don’t have to include all the. But, why would someone, why would you listen to music?” (P2)</td>
</tr>
<tr>
<td>Assuming</td>
<td>Assuming on behalf of the users</td>
<td>“How we will measure calories in it. Will user write?” “Maybe he will write” “Okay, I suppose” (P9)</td>
</tr>
<tr>
<td></td>
<td>Assuming relative importance</td>
<td>“Now, what is the most important thing to the user, to know the amount of calories or what he actually ate? I think the calories are the most important than other ones here” (P8)</td>
</tr>
<tr>
<td>Considering quality</td>
<td>Consistency</td>
<td>“Have a graph that tells the amount of workout and sleep” “That maintains consistency” (P9)</td>
</tr>
<tr>
<td></td>
<td>Usability</td>
<td>“I think this screen must be very easy and quick to use. It just can’t be a massive calendar” (P12)</td>
</tr>
<tr>
<td></td>
<td>Responsiveness</td>
<td>“Or you can make it like, responsiveness, more like on smaller screens to change the design, so I make it for that one.” “Maybe we could make it smaller” (P5)</td>
</tr>
</tbody>
</table>

### 3.5 Data collection and analysis

The sessions were transcribed by the first author. Although we did not correct participants’ grammar or malapropisms, we removed verbal static (e.g., “um”, “ah”, “uh”). We refer to each transcript by a unique identifier starting with a ‘P’. P1–P9 are the professionals; P10–P21 are the students.

We envisaged the data analysis in two phases: 1) inductive process coding to explore the cognitive mechanisms used by the participants; 2) deductive closed coding using concepts from existing literature to re-analyze the data through a specific theoretical lens. We used NVIVO (www.qsrinternational.com) to organize, analyze and visualize the qualitative data. The coding process is briefly described as follows:

#### 1) The first author performed process coding of all the transcripts (see Section 3.5.1).

#### 2) The second and the third author audited process coding.

#### 3) The auditing resulted in renaming some of the codes (e.g. doing design was renamed to making moves); some codes were combined together or rejected altogether (e.g. avoiding options and rejecting moves were combined to rejecting design moves).

#### 4) The first and the second author performed closed coding of one transcript together to ascertain the coding scheme (see Section 3.5.2).

#### 5) The first author then coded the rest of the transcripts.
6) The second and the third author audited closed coding.
7) The auditing resulted in minor changes such as adding new codes to the instances of fixation or critical thinking (e.g., rejecting design moves was added to fixation).

3.5.1 Process coding
We began by analyzing the data using inductive process coding [97]. That is, we coded each transcript line-by-line using gerunds (i.e., words ending with ‘-ing’). Each assigned label reflected the action contained in dialogues that shared similar characteristics (e.g., accepting requirements, discussing design moves). As the analysis progressed, some labels were reworded, subsumed by other similar labels or dropped. All codes that conveyed a particular process (i.e., action) were further categorized together to form themes, where a theme was seen as a high level conceptualization of action) were further categorized together to form themes, dropped. All codes that conveyed a particular process (i.e., action) were further categorized together to form themes, where a theme was seen as a high level conceptualization of multiple labels grouped together [98]. The saturation point was reached by the 14th transcript, i.e. no new labels or themes emerged on from the remaining seven transcripts.

3.5.2 Closed coding
Our previous experiments suggested the tendency of participants to get fixated on desiderata when presented as TRS. The effects of fixation can be minimized by a critical evaluation of the problem situation. Any attempt to critically evaluate the TRS should help participants to avoid fixating on the given TRS. Therefore, we applied an a priori coding scheme to compare instances of fixation against instances of critical thinking. Here, fixation refers to instances where participants: 1) accept aspects of the task (i.e., requirements, priority levels) without any discussion or reflection; 2) adopt properties of known examples without any discussion or reflection; or 3) reject, without any discussion or reflection, new ideas that diverge from given task structure or known examples. Critical thinking meanwhile refers to instances where participants critically evaluate or deviate from task parameters or known examples. In other words, if participants question something, but then accept it, we label it as critical thinking. We then counted these instances and compared.

4 RESULTS
This section presents the results from both analyses.

4.1 Process coding
Process coding produced seven themes, each of which we interpret as a distinct cognitive activity. Table 1 summarize the evidence for each themes, while the complete analysis is available in our replication package (see section 6).

4.1.1 Making design moves
A design move is a change to a design description [99]. Considering and making design moves was the participants’ most frequent activity. We found a total of 48 instances in fourteen groups where participants willingly tried to come up with multiple design ideas, reflected on those ideas and then made a move by selecting the most optimum one. However, participants in nine groups made design moves only intending to satisfy all the given requirements without assessing or reflecting on them. Out of these, six groups only tried to meet the requirements prioritized as high or high-medium.

4.1.2 Uncritically accepting
We observed participants uncritically accepting their initial design ideas and aspects of the task (e.g., requirements, priority levels) without any discussion or reflection. Participants were keen to adopt and force features of existing examples into their design concepts without assessing the existing designs. We observed imbalances in the pairs where Partner A would immediately accept Partner B’s ideas, while Partner B would unthinkingly reject Partner A’s ideas whenever they diverged from Partner B’s ideas.

4.1.3 Rejecting
Another pattern that emerged was the tendency of participants to explicitly reject any requirements or design ideas for various reasons (e.g., ambiguity, difficulty). Moreover, participants appeared reluctant to satisfy requirements prioritized as low or medium-low and requirements about which they had no knowledge. Participants would either temporarily ignore a high-priority requirement, or would permanently dismiss a requirement. When participants reject requirements, they often did so without any discussion, reflection or evaluation. Thirteen pairs explicitly rejected any idea or design move that diverged from their first design concept. We observed both uncritically accepting initial ideas (as discussed above) and uncritically rejecting new ideas that diverge from initial ideas.

4.1.4 Grouping
Fourteen of the pairs made sense of the desiderata by grouping requirements they perceived as similar. Out of these, eight groups were based on the priority levels provided, others on other sorts of similarity. For example, grouping given requirements as non-functional features or as pop-up notifications. Subsequent design moves appear to be informed by these groups. The tendency of participants to focus on grouping only high priority requirements while avoiding the low priority ones (we observed a total of 47 such instances) can be related to participants’ uncritical acceptance of priority levels and task presentation more generally.

4.1.5 Questioning
While participants often uncritically accepted aspects of the task (see section 4.1.2), they questioned others. Here, questioning refers to critically appraising something. Questioning is related to but distinct from rejecting. Sometimes participants questioned something before rejecting it; other times participants questioned something before accepting it; and other times participants rejected something without really questioning it first. Typically, one participant would raise doubts about a something (e.g., requirements, priority levels, existing examples). The pair would then discuss and come to a consensus about accepting or rejecting the concept. Participants also backtracked on their earlier design moves based on their evolving understanding of the task.
4.1.6 Assuming

Eleven pairs made explicit assumptions about the task. These assumptions were basically cognitive shortcuts that participants used to help them create designs easily with minimal information processing. We consider these assumptions as a deviation from the real or the observable facts. Participants appeared to speculate and derive at rather specific conclusions about how they perceived the requirements, instead of making actual sense of the problem situation. Five groups unreasonably perceived high priority requirements as more important than other low priority ones (e.g., counting calories as more important than recommending workouts); while other groups would over-estimate the effort required by creating self-imposed time-constraints. Participants also assumed a requirement (e.g., workout recommendation) was non-functional and jumped to conclusions about the complexity of initial design ideas.

4.1.7 Considering quality criteria

While planning the designs, participants would often express the need for certain aesthetic qualities for their solution designs. Participants in nine groups explicitly tried to change or alter their design moves for various quality criteria including usability, consistency of user experience, system responsiveness, speed, stability and aesthetics.

4.2 Closed Coding

This section presents the results of our closed coding. We identified 1006 instances of fixation compared to 298 instances of critical thinking. Table 2 presents the list of labels classified in each category and the corresponding number of instances of each label. Below, we briefly discuss each category and interpret our findings.

4.2.1 Fixation

All of the participants showed a tendency to agree and to accept instantaneously aspects of the task. We found 178 instances of participants accepting requirements without question and 145 cases of accepting priority levels without question. Participants appeared to accept task structure without any discussion of or reflection on the importance or the validity of the given desiderata. We also found 41 instances where pairs explicitly rejected new design ideas because they diverged from the initial design ideas.

Moreover, we found 517 instances where participants expressed complete confidence and extensively favored their initial (i.e., early) ideas by avoiding any speculation, discussion or reflection. We observed 79 cases across 19 pairs where participants tried to conceptualize their solution designs based on either a successful example or their previous experience. For example, “And also, kind of integration with Spotify or, another provider like Pandora like in other health fitness app I have come across” “We should do exactly that” (P1).

In 22 instances, participants said or implied that the system should satisfy all of the requirements; in ten instances participants said or implied that the system should satisfy at least all of the high and medium–high priority requirements. This is surprising because the requirements are intentionally dubious (as explained in Section 3.4).

4.2.2 Critical thinking

We found substantially fewer instances where participants attempted to think about the task parameters critically. In 199 instances pairs critically discussed design moves. However, these discussions were mainly about planning and organizing the way the application would look, and less about selecting the best option from multiple design options or assessing the importance of the given requirements. We observed eleven instances in six pairs of backtracking on their earlier design decisions. Such instances were characterized by participants reflecting on their initial design decision, followed by discussing alternative ideas and then selecting the one perceived as most appropriate. Some participants explicitly tried to generate multiple design options. We found very few instances where participants critiqued or reflected on the specific aspects of the task itself. Six pairs questioned priority levels with 14 instances; only two pairs expressed overall doubts about the requirements with four instances.

Similarly, we found nine instances of participants questioning the available technology and six instances of challenging existing examples. (Note: we did not provide examples; participants looked up examples on their own during the study.) Only three groups attempted to generate ideas that were not evident from the task materials, found in similar applications or generated by other groups. Only one pair expressed the need to consult the client to clarify specific ambiguous requirements.

4.2.3 Differences between students and professionals

Table 3 summarizes the differences between the professionals and the students. Professionals had more instances of both fixation and critical thinking than students; however, these differences are not statistically significant (Independent samples t-tests with effect size via Cohen’s d; fixation: t = 0.49, p = 0.63, d = 0.23 ± 0.87; critical thinking: t = 0.91, p = 0.37, d = 0.42 ± 0.89). Moreover, the ratio of fixation to critical thinking is almost identical (0.76 professionals vs. 0.78 for students). In other words, both students and professionals seem about equally susceptible to fixation.

Surprisingly, months of development experience is positively correlated with fixation (Pearson correlation; r = 0.528; p = 0.014)—see Fig. 2—but uncorrelated with critical thinking (r = −0.93; p = 0.689). In other words, more experienced developers were more prone to fixation. While these are post hoc tests on convenience samples, the results question the idea that fixation is limited to amateurs or that experience naturally mitigates it.

5 DISCUSSION AND IMPLICATIONS

5.1 Theory

Previous work [9], [10], [31] showed that presenting desiderata as TRS diminished creativity, which undermines SE success. The purpose of this study is to investigate how presenting desiderata as TRS diminishes creativity; that is, the

1. $F/(F + C)$ where $F$ is the mean number of actions associated with fixation and $C$ is the mean number of actions associated with critical thinking.
TABLE 2
Closed coding analysis

Pairs | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | Total
**Category 1: Fixation**
Accepting early ideas | 37 | 11 | 21 | 10 | 21 | 10 | 42 | 50 | 34 | 19 | 24 | 32 | 17 | 22 | 27 | 16 | 39 | 14 | 24 | 31 | 16 | 517
Accepting reqs | 10 | 2 | 4 | 5 | 10 | 12 | 5 | 3 | 7 | 6 | 14 | 8 | 23 | 8 | 7 | 12 | 15 | 7 | 8 | 5 | 7 | 178
Accepting priorities | 12 | 8 | 5 | 3 | 9 | 7 | 11 | 7 | 2 | 9 | 5 | 8 | 7 | 4 | 6 | 7 | 9 | 9 | 11 | 2 | 4 | 145
Accepting features of existing examples | 3 | 6 | 12 | - | 3 | 2 | 2 | 5 | 3 | 3 | 4 | 7 | 3 | 5 | 4 | 2 | 4 | - | 2 | 3 | 6 | 79
Rejecting design moves | 2 | 5 | 4 | 2 | 7 | - | 4 | 2 | - | - | 2 | 1 | - | 2 | - | - | 2 | 6 | 2 | 41
Trying to satisfy all reqs | 1 | 3 | 2 | 1 | - | 3 | - | 4 | 2 | - | - | 2 | - | - | - | - | 4 | - | - | 22
One participant accepts others decision | - | - | - | - | 5 | - | - | - | - | - | - | 3 | - | - | - | - | 6 | - | - | 14
Trying to satisfy high priority reqs | 2 | 1 | - | - | 1 | - | 1 | - | 2 | - | - | - | 1 | - | 1 | - | 1 | - | 1 | 10
**Total instances of fixation** | 67 | 36 | 48 | 21 | 43 | 47 | 60 | 74 | 50 | 39 | 47 | 55 | 57 | 40 | 44 | 40 | 67 | 31 | 57 | 47 | 1006

**Category 2: Critical thinking**
Discussing design moves | 4 | 15 | 8 | 2 | 9 | 9 | 4 | 16 | 13 | 12 | 19 | 7 | 5 | 12 | 17 | 3 | 11 | 14 | 5 | 8 | 6 | 199
Generating multiple design options | - | - | 2 | 1 | 2 | 1 | 6 | 5 | 3 | - | - | 5 | - | - | - | 2 | 1 | - | 4 | 2 | 48
Questioning priority levels | - | - | 2 | - | - | - | - | - | 3 | - | - | - | - | 1 | - | - | - | 3 | - | - | 14
Backtracking on earlier decisions | - | - | - | 2 | - | 1 | 3 | - | 2 | - | - | - | - | 1 | - | - | - | 2 | 11
Questioning available technology | - | 2 | - | - | 1 | - | - | - | - | 1 | - | - | - | - | 1 | - | - | - | 2 | - | 9
Questioning existing examples | 2 | - | 2 | - | - | - | 1 | - | - | - | - | 1 | 1 | - | - | - | 1 | - | - | - | 6
Generating innovative ideas | - | 2 | - | - | 2 | - | - | - | - | 1 | - | - | - | - | 1 | - | - | - | 2 | - | 5
Questioning reqs | - | 2 | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 4
Questioning client’s needs | - | - | 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 2
**Total instances of critical thinking** | 6 | 30 | 21 | 5 | 14 | 12 | 12 | 25 | 17 | 14 | 22 | 12 | 6 | 14 | 18 | 5 | 14 | 20 | 5 | 13 | 13 | 298

TABLE 3
Differences between students and professionals

<table>
<thead>
<tr>
<th>MEAN</th>
<th>PROFESSIONALS</th>
<th>49.67</th>
<th>46.67</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN INST. OF CRITICAL-THINKING</td>
<td>15.87</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>RANGE (FIXATION)</td>
<td>21–74</td>
<td>31–67</td>
<td></td>
</tr>
<tr>
<td>RANGE (CRITICAL THINKING)</td>
<td>5–30</td>
<td>5–22</td>
<td></td>
</tr>
<tr>
<td>FIXATION RATIO</td>
<td>0.76</td>
<td>0.78</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Scatterplot of experience (months) vs. fixation (instances)

Fig. 3. Theoretical framework

Note: Boxes indicate constructs; arrows indicate causality; plus and minus signs indicate direction of effect.

cognitive mechanism mediating the previously established causal relationship. The results of this study suggest that providing designers with TRS induces requirements fixation and hinders critical thinking, thereby negatively affecting creativity (see [9], [10]); as shown in Fig. 3—see Table 4 for construct definitions.

The simulation reported above shows that participants given TRS have many more instances of fixation than critical thinking. Furthermore, we identified several indicative behaviours for both fixation and critical thinking—the labels in Table 2. For example, unthinking acceptance of TRS indicates requirements fixation; questioning the reasonableness of priority levels indicates critical thinking.

5.2 Implications

The interplay between fixation, critical thinking, creativity and the presentation of desiderata as TRS has numerous implications for SE professionals, researchers, and educators.

For professionals, the best way to present desiderata involves trade-offs among many criteria including clarity, understandability, flexibility, modifiability and, of course, creativity (100, 101). Where creativity is a priority,
avoid TRS and over-structuring, over-simplifying and over-rationalizing problem statements. Try to present desiderata in ways that encourage skepticism and critical thinking. Information presented to the designers should ideally be less-structured and easily modifiable. In contrast, a clear, well-structured TRS might help high correctness of code, which can increase the possibility of success of a software development for mission- or safety-critical domains.

For researchers, it is critical to abandon the naive view that analysts elicit requirements and that design transforms them into appropriate system features. This view obscures the actual relationship between RE and design, which remains contested and poorly understood. SE occurs in complicated situations where stakeholders disagree on system goals and desired features. Analysts and users co-construct evanescent preferences rather than eliciting firm and robust requirements. Design is a creative, improvised, non-deductive process in which designers imagine new systems rather than rearrange old ideas. Expert designers in other fields resist initial problem frames and solution conjectures; they do not deliver requirements in a box-checking manner. Designers need to elicit requirements from stakeholders, to discover meaningful and valid requirements that are not directly observable. Reality is only imperfectly and probabilistically apprehensible. Rather than discovering causal laws, scientists therefore construct explanations based on “generative mechanisms”—the powers objects have to influence each other.

Making concrete recommendations for SE education is more difficult. Obviously, courses that present an outdated, positivist view of RE should be updated. Non-empirical legacy concepts such as the waterfall model and project triangle should be replaced with evidence-based concepts and theories. Beyond that, we want to recommend teaching a host of underrepresented subjects including design thinking, creativity techniques and theories of cognitive biases. However, SE curricula are already tight. Perhaps a more tractable approach is to transition students to less and less structured assignments as they advance. More open-ended assignments with ambiguous goals, conflicting stakeholder preferences, ill-structured problems and incomplete specifications should help prepare students for more realistic software contexts.

5.3 Quality criteria and threats to validity

We see protocol studies as most consistent with critical realism. Critical realism is a body of philosophical work that attempts to solve Hume’s problem of induction by merging a realist ontology with a relativist ontology. Critical realism is fundamentally different from both positivism and constructivism. Positivism (and falsificationism) view reality as observer-independent, objective, measurable and characterized by universal, deterministic, counterfactual, causal laws. Constructivism, meanwhile, views reality as observer-dependent, subjective and devoid of universal laws because knowledge is context-dependent. Positivism embraces (epistemological) realism; Constructivism embraces epistemological and ontological relativism.

In contrast, critical realism blends a realist ontology (“transcendental realism”) with a relativist epistemology (“critical naturalism”). In other words, critical realism assumes that the phenomena that scientists study are real whether they can be directly observed (e.g., people, length, Mars) or not (e.g., electrons, creativity, quasars). But because social reality resists experimental closure and is rich in unobservable properties, reality is only imperfectly and “probabilistically apprehensible”. Rather than discovering causal laws, scientists therefore construct explanations based on “generative mechanisms”—the powers objects have to influence each other.

Since critical realism is fundamentally different from positivism and constructivism, it has different evaluation criteria, namely—“ontological appropriateness”, “contingent validity”, multivocality, “trustworthiness”, “analytic generalization” and “construct validity”. These criteria do not map neatly into either positivist criteria (internal validity, external validity, etc.) or constructivist criteria (credibility, transferability, etc.)

Critical realism is ontologically appropriate because the whole point of a protocol study is to explore cognitive phenomena that are real but cannot be observed directly. Contingent validity is the degree to which the study explores generative mechanisms rather than deterministic causal laws. Again, here we are explicitly concerned with exploring the generative mechanism that accounts for design creativity.

Multivocality—the degree to which research integrates diverse perspectives—is typically achieved through data triangulation, which is difficult in a protocol study. Our dialogue-based approach allows us to examine the statements of each half of a participant-pair, as well as directly observing and comparing pairs. However, we cannot corroborate our findings against independent data sources such as archival records, like in a case study. Moreover, brain scans are not yet sophisticated enough to cross-check the inferences we make from participant’s verbalizations.

Trustworthiness refers to the chain of evidence from observations to conclusions (see Tables 1 and 2 and the ability of an independent researcher to audit or replicate the findings. We provide as much detail of our analysis process as possible within space limitations, and all of the materials necessary to run an identical study with new participants.
However, for privacy reasons, we cannot publish the full transcripts of the design sessions and therefore an independent researcher cannot directly audit our coding.

As we refined the conceptualization of themes, we often renamed or merged multiple themes and their corresponding labels. For example, the theme expressing values was renamed to considering quality criteria and merged with an earlier theme non-functional requirements, and the labels discussing alternative ideas was renamed to discussing design plans. Despite much refining of labels, the themes and their relationships stabilized early and remained stable. The frequencies of the labels (i.e., fixation and critical thinking) in Table 2 are unweighted and do not provide evidence of the total number of ideas gained or lost due to one instance of critical thinking and fixation respectively. In other words, one instance of fixation might be more or less important than one instance of critical thinking for creativity.

Moreover, since creativity is only one of the many antecedents of SE project success [24], less constrained and restrictive presentations of desiderata may also undermine success through some other mechanisms, e.g., legal or mission/safety critical constraints. Furthermore, TRS could affect creativity through mechanisms other than those considered in this study.

Analytic generalization refers to generalizing from observations to theory, rather than from a sample to a population. We generalize from observations of designers to the theory shown in Fig. 3.

A protocol study is non-statistical, non-sampling research, using a convenience sample of participants completing a particular task in a particular environment. Our participants were mostly male, non-native English speakers working in English, completing a single, artificial task, in an unfamiliar design domain, in an artificial environment, using artificial task materials, while being watched. All of these factors may have affected our results in unknown ways. Results cannot be statistically generalized to different people, other tasks, other environments, or other ways of representing desiderata (e.g. user stories, goal models).

Construct validity refers to the degree to which the operationalization and measurement of constructs supports scientific inferences. The only constructs in this study are fixation and critical thinking, which we operationalize by having an expert judge identify them. However, this labeling process is intrinsically subjective, so another analyst might label the data differently. We mitigated this threat by having the second and third authors review the first author’s coding, leading to numerous revisions and clarifications. However, a different research team might still produce different labeling.

5.4 Future research directions

We see several promising avenues for future work:

1) Experimentally comparing different representations (e.g., user stories, use cases, code) of the same desiderata to determine which representation is most effective to foster creativity in different circumstances;

2) Creating techniques, tools and practices for modelling and managing ambiguity and conflict; and

3) Using eye-tracking or protocol analysis to study what professionals attend to and ignore while designing software.

Moreover, requirements fixation is just one of several cognitive biases that may hamper creativity in software design. Future work should investigate related cognitive phenomena including:

1) Confirmation bias—attending disproportionately to information that confirms our current beliefs [22].

2) Miserly information processing—the tendency to avoid deep or complex information processing [85].

3) Conceptual fixation—considering only one or a small number of solution concepts [35].

4) Design fixation—sticking too closely to given or known examples [105].

While confirmation bias, miserly information processing, conceptual fixation and design fixation have all been studied extensively, little work has investigated their effects on software design in particular [106].

6 Conclusion

In summary, desiderata are things that project stakeholders prefer, want or need in a software system. Desiderata can be presented in many ways (e.g. templated requirements specifications, user stories). Previous research showed that presenting desiderata as templated requirements specifications led to less creative designs. We therefore conducted a dialog-based protocol analysis to investigate the cognitive mechanism by which templated specifications affect design creativity. We analyzed the data in two ways: inductive process coding and closed coding. Process coding revealed seven kinds of design actions: making design moves, uncritically accepting, rejecting, grouping, questioning, assuming and considering quality criteria. Closed coding showed that actions associated with requirements fixation are significantly more frequent than actions associated with critical thinking.

These results suggest that presenting desiderata more restrictively as templated requirements specifications is associated with less critical evaluation of task structure and less critical thinking. In other words, templated requirements specifications inhibit design creativity because designers get fixated on desiderata presented (i.e. written) restrictively, well-structured and constrained language, hindering critical thinking.

This paper therefore makes three main contributions: (1) it advances a theory that explains the (previously established) relationship between templated requirements specifications and design creativity; (2) it elaborates the concept of requirements fixation; (3) it presents a simple taxonomy of software design actions. However, our results do not indicate that requirements analysis is useless or that more analysis is counterproductive to creativity. The paper just attempts to present the underlying cognitive mechanisms explaining the effects of presenting desiderata in a very specific way—as templated requirements specifications—on design creativity.
While previous experimental research has demonstrated that presenting desiderata as templated requirements specifications reduces design creativity, our current research explores the underlying cognitive mechanisms that explain this relationship. The results of this study indicate that, given templated requirements specifications, software designers do not proceed as we might hope. Designers should carefully evaluate each desideratum before accepting or rejecting it for articulable reasons. Our observations suggest that designers tend neither to critically evaluate requirements nor to reject questionable ones.

7 DATA AVAILABILITY
A comprehensive replication package including all the task documents (i.e., a list of prioritized TRS, demographic questionnaire and blank design template) and the results of the process coding analysis with example quotes are stored in the Zenodo open data archive [17]. (Note: we do not include the transcribed recordings in the replication package to maintain the anonymity of the participants).

ACKNOWLEDGMENT
This study was partially supported by the HPY: Research Foundation (HPY:n Tutkimussäätiö Apurahat) grant.

REFERENCES


Rahul Mohanani, PhD (Oulu University), MSc (Lancaster University), B.Eng. (Mumbai University), is a senior scientist at Fortiss GmbH, Munich and an Adjunct Asst. Professor of Software Engineering at IIT Delhi. His research, intersecting empirical SE, human aspects and design thinking, has been published in premier venues including the ACM/IEEE International Conference on Software Engineering and IEEE Transactions on Software Engineering. For more information, please visit [http://rahulmohanani.net](http://rahulmohanani.net).

Paul Ralph, PhD (British Columbia), B.Sc / B.Comm (Memorial), is an award-winning scientist, author, consultant and Professor of Software Engineering at Dalhousie University. His research intersects empirical software engineering, human-computer interaction and project management. Paul is a member of the IEEE Transactions on Software Engineering review board and chair of the ACM Paper and Peer Review Quality Task Force. For more information, please visit [https://paulralph.name](https://paulralph.name).

Burak Turhan, PhD (Boğaziçi University), is a Professor of Software Engineering at the University of Oulu and an Adjunct Professor (Research) in the Faculty of IT at Monash University. His research focuses on empirical software engineering, software analytics, quality assurance and testing, human factors, and (agile) development processes. He is a Senior Associate Editor of Journal of Systems and Software, an Associate Editor of ACM Transactions on Software Engineering and Methodology and Automated Software Engineering, an Editorial Board Member of Empirical Software Engineering, Information and Software Technology, and Software Quality Journal, and a Senior Member of ACM and IEEE. For more information, please visit [https://turhanb.net](https://turhanb.net).

Vladimir Mandić, PhD (Oulu University), is an assistant professor of SE at University of Novi Sad, Serbia. He received his PhD degree in Information Processing Science and SE from the University of Oulu, Finland, and M.Sc.E.E from the University of Novi Sad, Serbia. His areas of interest are software process improvement, empirical software engineering, goal-driven measurement approaches, technical debt and value-based software engineering. He is a member of the IEEE Computer Society.