2018 ACM/IEEE International Conference on Technical Debt

Trade-off Decisions Across Time in Technical Debt Management: A Systematic Literature Review

Christoph Becker
Faculty of Information, University of Toronto
Toronto, Canada
christoph.becker@utoronto.ca

Ruzanna Chitchyan
Department of Computer Science, Univeristy of Bristol
Bristol, UK
r.chitchyan@bristol.ac.uk

Stefanie Betz
Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology
Karlsruhe, Germany
stefanie.betz@kit.edu

Curtis McCord
Faculty of Information, University of Toronto
Toronto, Canada
curtis.mccord@mail.utoronto.ca

ABSTRACT

Technical Debt arises from decisions that favour short-term outcomes at the cost of longer-term disadvantages. They may be taken knowingly or based on missing or incomplete awareness of the costs; they are taken in different roles, situations, stages and ways. Whatever technical or business factor motivate such decisions, they always imply a trade-off in time, a 'now vs. later'. How exactly are such decisions made, and how have they been studied?

This paper analyzes how decisions on technical debt are studied in software engineering via a systematic literature review. It examines the presently published Software Engineering research on Technical Debt, with a particular focus on decisions involving *time*. The findings reveal surprising gaps in published work on empirical research in decision making. We observe that research has rarely studied *how* decisions are made, even in papers that focus on the decision process. Instead, most attention is focused on engineering measures and feeding them into an idealized decision making process. These findings lead to a set of recommendations for future empirical research on Technical Debt.

CCS CONCEPTS

• **Software and its engineering → Software design tradeoffs**; *Software evolution*;

KEYWORDS

technical debt, decision making, time, intertemporal choice, naturalistic, rationalistic, behavioral software engineering

ACM Reference Format:

Christoph Becker, Ruzanna Chitchyan, Stefanie Betz, and Curtis McCord. 2018. Trade-off Decisions Across Time in Technical Debt Management: A Systematic Literature Review. In *TechDebt '18: TechDebt '18: International Conference on Technical Debt , May 27–28, 2018, Gothenburg, Sweden,* ?? (Ed.). ACM, New York, NY, USA, Article ???, 10 pages. https://doi.org/10.1145/3194164.3194171

TechDebt '18, May 27–28, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5713-5/18/05…$15.00
https://doi.org/10.1145/3194164.3194171

1 INTRODUCTION

Technical Debt (TD) arises from decisions that are 'expedient'[26] in the short term but cause a need for possibly complex and costly actions in the medium or long run[1]. Technical debt thus always implies the notion of time, decisions, and trade-offs. These can be direct decisions about incurring debt, such as using an old platform version when a new version is already out; or indirect decisions that inadvertently cause so-called unintended debt, such as adding excessive code to one class without noticing the bad 'code smells' this causes. In this paper, we focus on decisions directly related to TD and its management.

The question arises: How do software professionals perceive this trade-off in time when making technical debt-related decisions? How exactly do they make such decisions on technical debt? And how does research study these questions?

This literature review analyzes how the literature on TD and TD management conceptualizes and studies the making of decisions related to TD by people in software projects when those decisions involve intertemporal choices. It takes inspiration from a distinction that is made in decision making research between normative approaches based on a prescriptive view on how decisions *should be made*, and descriptive approaches that focus on how *people make decisions*. Normative rationalistic theories are based on concepts of utility, rational choice and bounded rationality. Descriptive naturalistic decision making theories regards these as inappropriate and instead aims to understand decision making as it takes place in realistic environments. It is currently receiving increased attention, with the last Nobel prize awarded to behavioral economist Richard Thaler, who contributed significantly to expand decision making research beyond simplistic lab experiments. Our aim is to understand to what degree TD research has incorporated insights from naturalistic decision making research. The review's main contribution is to identify significant limitations in prior empirical work on decision making in technical debt, suggest concrete guidelines and frameworks to be considered in future empirical studies, and introduce the concepts and frameworks of 'intertemporal choice'[4, 10] to technical debt management.

2 BACKGROUND

The Software Engineering discipline has often utilized the work of other sciences to inform its perspective on decision making (e.g. [8]). *Behavioral Software Engineering* aims to place behavioral perspectives center stage and introduce relevant insights from behavioral sciences [19]. Researchers in economics, behavioral economics, psychology and neuroeconomics have studied decision making process for some time and brought forward a number of theories [10, 23] that could contribute to a better understanding of *intertemporal choices*, i.e. "decisions involving tradeoffs among costs and benefits occurring at different times" [10, :351]. A key distinction in decision making theories and approaches is between those that broadly assume a *rationalistic* stance and derive from a normative perspective, and those that have abandoned these perspectives in the face of contrary empirical evidence to follow what they called a *naturalistic* approach to decision making research that aims to examine real decisions made outside the confines of controlled lab experiments. The overwhelming majority of research on decision making across multiple disciplines such as behavioral economics and psychology has traditionally focused on *normative* and rationalistic approaches [24]. In this work, it is presumed that decision makers are above all rational agents who always identify and take 'the best' possible decision, where what is best is defined by an objective value function, building on game theory[36] and utility theory going back to Bernoulli. Models for such a decision making process are based on a formal value function that rational economic agents would apply to choose the best out of a number of possible decision options using a set of criteria[13, 32, 37]. This is the core of Multi-Criteria Decision Analysis (MCDA) [13]. It is worth highlighting that rationalistic decision making research in behavioral economics for the most part took place in the laboratory, using controlled experiments to study how subjects would go about well-defined and clearly delineated, but not fully realistic, choices. The assumptions of the theory are not empirical, but considered axiomatic. For decades, however, empirical research has pointed out that the assumptions of rationalistic decision theories contradict empirically observed evidence. For instance, people take different choices for an identical problem when its future outcome is stated in terms of gain (e.g., a 40% chance of future gain) as opposed to loss (e.g., a 60% chance of future loss), because gains and losses are valued differently [33], and this changes for future gains and losses [22]. Consistent empirical evidence suggests that normative theory, models, and assumptions about how people make decisions are often inadequate [2, 10, 23, 24].

Two responses to this emerged: The mainstream of research introduced adaptations to rationality to account for some of the observable inconsistency. Since humans are deviating from what these theories see as 'optimum', they are seen as only 'boundedly rational' in the work of Herbert Simon. Others focus on how human decision makers use 'heuristics' (shortcuts that come with biases) in their decision making, and how the framing of outcomes affects their decisions. Prospect theory [23, 35], for instance, suggests that decision making is based on the potential value of losses and gains to individual agents (rather than the final outcome) and the way that agents subjectively frame an outcome in their minds affects the expected outcome. Through this adaptation of the value function, prospect theory addresses some of the empirical inconsistencies of rational choice. It should be noted, however, that these theories still fall in the same paradigm: They are based on experiments conducted in a lab setting, with limited alternatives, and largely based on rationalistic assumptions that assume decision makers would enlist alternative options and decision making criteria, then compare their alternatives to select what seems the optimum decision. Arguably, these assumptions are inadequate[2].

As a result, the normative approach cannot account for the inconsistency of its model's assumptions with empirically observable behavior [2, 16]. For example, alternatives are assumed to be independent and preference relations transitive, but in practice, addition or removal of irrelevant alternatives can cause a reversal of preferences. In addition, researchers grew increasingly uncomfortable with the assumptions underpinning the lab experiment model as a sole representative of human decision making. To account for this, alternative, *naturalistic* approaches to study decision making in realistic contexts emerged [15, 16]. This is summarized in more detail in [4]. Most strikingly, their studies convincingly demonstrated that expert decision makers, even when trained in MCDA approaches, typically did not mentally enlist alternatives and criteria to guide their decisions, but employed a more sequential approach guided by their expertise and tacit knowledge[15, 17]. Instead of denouncing deviations from rationalistic theory as deficiencies to be overcome, researchers in this paradigm argued against the empirical validity of the underlying theory and suggested that a descriptive approach to studying real decision making by experts will provide a more realistic understanding of decision making in practice. Naturalistic research thus broke with the mainstream in two ways: First, it did not structure problems and observations in terms of rationalistic concepts such as utility, alternatives, value functions, and weighted criteria. Second, it abandoned the model of controlled lab experiments in favor of observations of real-world decision making, arguing that this will ultimately be able to provide more effective decision support [15, 21].

Given that the inadequacies of normative models are becoming well recognised, we could expect the Software Engineering profession to widely adopt the naturalistic decision making models, especially in field studies. Yet, a recent literature review on time and trade-off decision making in Software Engineering [4, 5] observes a continued dominance of normative decision making mindset in our discipline. In the present study, we aim to investigate what view is held on decision making within the SE technical debt community.

By definition, TD decisions involve intertemporal choices, albeit that choice can often be implicit, rather than clearly articulated. As this topic area is specifically concerned with trade-offs across time, one would expect that the above discussed inadequacies of the normative models would be both recognised and addressed in the SE work that concerns technical debt. Yet, no previous work has studied this issue. Thus, in this paper we set out to explore *how technical debt and technical debt management conceptualizes and studies the making of decisions by people in software projects.*

The overall aim of this study is to **examine the assumptions underpinning intertemporal choices** regarding technical debt and its management. To accomplish this we turn to the literature published on this topic within software engineering and project management communities, and:

(1) systematically review and analyze the publications corpus on technical debt and map out its key groups of papers;
(2) identify the assumptions and decision making theories underpinning the study of decision making in this area;
(3) articulate opportunities for future research on this subject.

Through a systematic literature review process (presented in Section 3) we have chosen and analyzed a corpus of 240 papers. We discuss (in Section 4) how empirical studies conceptualize and present time and trade-off alternatives within technical debt management literature, identify gaps in how decision making is reviewed and investigated, and map how SE literature approaches trade-off decision making over time with respect to technical debt. Section 5 compares this to related work, and Section 6 summarizes key findings and infers recommendations for future research.

3 STUDY DESIGN

Since the present paper is principally focused on the study of the assumptions in decision making for technical debt, we select papers that report on empirical evidence, since these are the studies that examine *actual decisions*. We leave out the detailed study of non-empirical reports as future work.

The specific *Research Questions* formulated to help achieve the aim of our study are:

RQ1 Which **research** methods have been used to study ***technical* debt trade-off decisions** involving time? Here we are more specifically interested in empirical vs. other methods.
RQ2 Which **assumptions** on decision making underpin empirical studies of such decisions?
RQ3 What role does **time** play in these trade-off decisions?

3.1 Search Strategy

The overall search process for this literature review is based on guidelines for systematic literature review (SLR) established by Kitchenham [14] (though the corpus of papers was initially elicited for a related study[1], as reported in section 5). The process is summarised in Figure 1.

The *search string* - "technical debt" - was conceivably straightforward, since this term is very well-established in SE. Technical debt is defined as: "a design or construction approach that is expedient in the short term but that creates a technical context in which the same work will cost more to do later than it would cost to do now (including "increased cost over time" [9, borrowing from [26]]).

We found 620 papers. The resulting corpus of papers was revised applying the selection criteria presented in section 3.2, after which 240 papers remained for analysis.

Information sources used for this study are the indexing systems and digital libraries most commonly used for publishing software engineering research: Scopus, IEEE Xplore, and ACM Digital Library. An automated search was performed over these sources using the search string.

[1]The corpus was initially compiled through an adjacent search in the preceding literature review [5], which aimed to investigate how trade-off decisions (over time) have been conceptualized within SE literature using a broader search strategy.

3.2 Selection Criteria

We established the following criteria to identify relevant publications:

- *Publication Year*: All years were included.
- *Publication Type*: Included peer-reviewed papers published in journals, conference proceedings, and workshop proceedings.
- *Content*: Papers had to contain discussion of decision-making *in* software engineering projects.
- *Coverage*: Papers had to cover development of software systems rather than only hardware.

We excluded papers that were:

- Published in languages other than English;
- Retracted by the publisher;
- "Non-paper" results such as: posters, abstract-only submissions, book reviews, books, entire volumes of proceedings (note: matched individual papers from volumes were included), panels, presentations, tutorials, short opinion pieces.
- PDF was unavailable (e.g., behind a paywall or not locatable).

The 240 papers that met the selection criteria were then analyzed as per the procedures and methods discussed below.

3.3 Data Extraction and Analysis

For data extraction and analysis, we used a qualitative content analysis method [25, 28]. The data obtained here allowed us to address RQ1.

At the 2nd stage a more in-depth qualitative text analysis was carried out for the sub-category of papers (identified at the first stage of analysis) that presented empirical studies of decision making. This analysis allowed us to address RQ2 and RQ3.

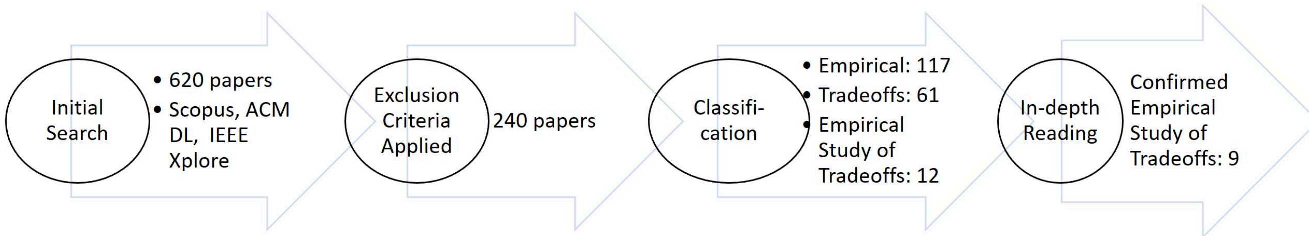
Data extraction and analysis was carried out by the three authors of this paper, who reviewed and coded each paper. The initial set of codes was defined in accordance with the categories deemed relevant in the study design phase. A code-set from the previous related study [5] that focused on decision making in Software Engineering was used as starting point and updated through the coding activity. The initial codebook, as well as the updates, were discussed and agreed upon by all collaborators. A web-based spreadsheet was used to support the coding and review process. Free annotation was also used to capture additional information the coders deemed relevant. The final codebook and coding data are published [3].

These coding stages are detailed below.

Stage 1– Mapping research methods and focus. For this stage, each of the author was assigned a third of the papers as a primary coder. Additionally, the review load of each primary coder was split equally between the other two authors for secondary coding. The aim of the secondary coding was to ensure that each coder's work is cross-validated equally by two other colleagues. For this, the secondary coders independently coded every fourth paper from the set assigned to them, resulting in 74 double-coded papers (30.8%).

The coding was conducted in five rounds, each followed by a discussion between the coders. Any disagreements and inconsistencies in primary and secondary coders' views that arose through this process were discussed until consensus was reached. Papers of which disagreements arose in the first round were then recoded by

Figure 1: The study process



both researchers who initially coded them. In a few cases where the two coders were unable to agree, the third author was asked to review the paper and a three-way discussion was used to agree on code categories. Disagreements fell markedly after initial discussion, so that the coders had high confidence in the reliability of the remaining codes.

At this stage the publications were classified according to the:

- **type** (using types from [39]: technical solution, research evaluation, validation, or philosophy, opinion research, personal experience),
- **scope** of the decision studied (i.e., decisions in project management, requirements engineering, architecture, maintenance, other),
- the **methods** of investigation and research (i.e., empirical study of decision, empirical study but not of decisions, literature review, other),
- the **focus** of empirical research (real-world decisions, experiments and quasi-experiments, focus on people, other).
- whether there was a **trade-off** decision (yes/no question), and if so,
- the **dimensions** considered (e.g. cost, functionality, quality, time, risk, maintenance, and others), and finally,
- the **assumptions** about decision making theories that could be identified (normative, descriptive, both, unclear).

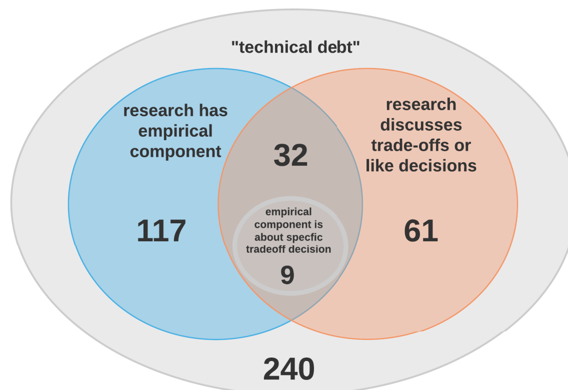
Stage 2– In-depth Content Analysis. For the in-depth analysis each of the selected empirical papers was assigned to one primary and one secondary coder, so each article was coded by at least two reviewers independently.[2]

These papers were analyzed with respect to the decisions that they studied, decision models and assumptions that they carried, aspects they considered, and trade-offs accounted for in the decision process. Since the decision models and assumptions are often implicit in the publications and often spread across various parts of the articles, we used a more flexible process of iterative qualitative coding.

Here too, the initial code set was derived from a previous, related study [5], but was flexibly updated to focus on how the papers characterized and studied decision making. For coding decision making theories and assumptions, the agreed coding guidelines required each coder to clarify for each paper whether the identified decision making theories are explicitly stated or implicitly derived; include detailed explanations and quotes from the text to substantiate the coding; and highlight specific assumptions and theories they identified (expected utility, discount factors, rational choice, bounded rationality, Multi-Criteria Decision Analysis, ...) as well as explicit methods such as the Analytic Hierarchy Process or Multi-Attribute Utility Analysis.

[2]The first author coded all papers. Thus, the paper where he was not assigned as 1st or 2nd reviewer were coded by three authors.

Figure 2: Segments distinguished according to research method with a focus on decisions



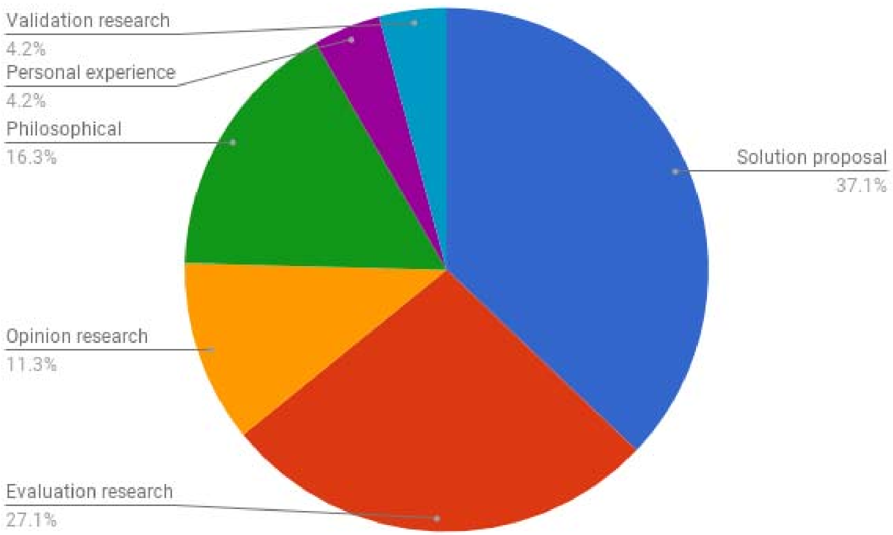
4 FINDINGS

Below we report on the findings from our categorisation and detailed analysis work conducted for the technical debt and its' time-related aspect in the decision making literature published in Software Engineering domain. We address each of the previously set research questions individually.

4.1 RQ1: Research methods used to study TD trade-off decisions involving time

As noted above, the publications analysed in this study were coded for the empirical focus on studying technical debt related decisions, and for including, in particular, decisions involving time. Out of the 240 papers included into the final set for analysis, we found that 117 papers had some degree of empirical component, and 61 discussed trade-offs in such capacities. The resulting Venn diagram, shown in Figure 2, demonstrates a quite even distribution across the emerging sub-segments, but indicates that **only 9 studies** explicitly used *empirical methods* to study specific *trade-off decisions*. This set represents papers that attempt to examine decision-making in software engineering in real or experimental situations. This

Figure 3: Allocation of the papers according to Wieringa's classification [39]



number is consistent with (but even lower than) the similarly low number of studies with empirical decision making focus in Software Engineering reported in a related study [5].

Furthermore, we observe that the majority (over 37%) of papers that discussed trade-offs in technical debt are focused on solution delivery, as shown in Figure 3 (using classification of research types suggested by Wierenga at al) [39]. While a good proportion of papers (27.1%) is tackling research evaluation, only 16 papers consider trade offs in the evaluation setting. As noted above, only 9 of these evaluate the results of actual empirical studies. The set of these 9 studies is described in Table 1 below.

4.2 RQ2: Assumptions that underpin decision making

Normative perspectives of rationalistic decision making dominate all empirical papers. While this is usually not made explicit (similar to the findings in [5]), the overall assumptions that surface in these studies are:

(1) Decisions are made by evaluating and weighing multiple alternatives against multiple criteria, and
(2) Obtaining measures for each criterion is seen as the central research focus.

As one article states, "The goal of identifying and measuring technical debt is to facilitate decision making." [12, :166] However, how those decisions really are made is rarely studied directly in these articles, even in those that focus on the decision process. Instead, most attention is focused on engineering measures and feeding them into an idealized decision making process. The discussion in Guo [12] exemplifies this: "In theory, any decision-making criteria could have been incorporated into the process of estimating principal and interest ... If the personnel best suited to paying off a TD instance are not available, then this could be incorporated into that instance's estimate of principal, making the principal high enough that it would not be considered cost-effective to pay off." Here the decision making theory is not explicitly named, but it is clearly normative rationalistic decision making that maximizes utility expressed in terms of monetary value.

Even though in some studies, empirical evidence suggests that the decision makers deviated significantly from the proposed normative decision making methods, no study explicitly questioned the normative rationalistic paradigm. For example, the concern classification scheme used by Guo et al. [12] is restricted to "costs (initial overhead, other costs), benefits (observed benefit, perceived usefulness), obstacles (difficulty in TD identification, Time pressure, Process integration) and Process deviation (TD estimation, Task prioritization)." These highlight exclusively those factors that arise *if* normative rational choice is assumed to be an adequate description of the studied decisions.

Other studies are even more narrowly predicated on MCDA methods. For example, Snipes asked their participants "what factors are considered when you make a decision about when to fix a defect?" and "how are these factors weighted?" [31]. The study design did not allow for the consideration that the decision makers may not weight criteria at all, as suggested by naturalistic models such as Recognition Primed Decision making [17].

Similar to studies in other fields, empirically observed decision making differed from the proposed normative models. In Guo [12], the authors state that "the criteria used for decision making about TD differed between the proposed TD management approach and the actual projects." The data categorisation scheme for this study was predicated on the notion of multiple criteria decision analysis methods. Consequently, one could only wonder as to what 'deviation' (beyond the criteria themselves) would have been observed if the possibility of alternative decision making models had been included into the data analysis process. Consider this description of a specific deviation: *"For test automation, they did not do numeric estimation for the principal and interest as our approach requires. A*

Table 1: Empirical studies characteristics

Title (abridged)	Year	Authors [reference]	Research Method and Summary	Cited (GS)
Toward Design Decisions to Enable Deployability: Empirical Study of 3 Projects	2014	Bellomo et al. [6]	An **interview study** with three project teams is conducted that focuses on deployability goals and architectural design decisions. The key focus lies not on those decisions in particular, but on the extraction of useful tactics and patterns as a contribution to continuous integration.	24
Costs and obstacles encountered in technical debt management	2016	Guo et al. [11]	The article describes a **case study** in which a project attempts to apply a TD management approach developed by the researchers. It discusses how the application deviated from the method, presents identified reasons and obstacles, and evaluates the costs of TD management.	1
Exploring the costs of technical debt management: a case study	2016	Guo et al. [12]	A **case study** is conducted in which a project attempts to apply a TD management approach developed by the researchers (similar to above, but a different study). The article describes how the application deviated from the method, discusses possible reasons, and aims to quantify the costs of TD management.	28
An Enterprise Perspective on Technical Debt	2011	Klinger et al. [18]	A small **interview study** is conducted with four architects at one company, focused on retrospectively understanding how their teams made decisions to incur TD. Questions focused on the nature of technical debt and its context, the benefits of TD, decisions to repay it, and factors in the decision making. The aim was also to explore to what degree TD can be seen as a financial tool.	60
Decision-Making Framework for Refactoring	2015	Leppanen et al. [20]	This is a **multiple case study** focused on identification of stages in refactoring, and analysis of factors leading to refactoring decisions and triggering the transition between stages. For this interviews are conducted with 3 employees of 3 companies. Then software system from one of the companies is examined to compare findings from a specific project to those reported through interviews.	2
Managing Technical Debt in Software Projects Using Scrum: An Action Research	2015	Oliveira, Goldman, Santos [27]	This is an **action research** study, where authors select a particular TD assessment framework and applied,it within the settings of two company's,projects with Scrum process. Data is collected for each stage of the framework's application. The authors observe that each role could bring about a role-specific debt (e.g., usability debt, documentation debt), and that parts of the debt	2
Managing Technical Debt in Practice: An Industrial Report	2012	Siebra et al. [30]	The paper presents a **case study** on decision making within an industrial project. The project's data was collected from emails, documents, CVS logs, code files and interviews with developers and project managers for the full period of the project's active live (6 years). The data is analysed, with a set of factors identified that influence the project's decisions through time.	17
Defining the Decision Factors for Managing Defects: A Technical Debt Perspective	2012	Snipes et al. [31]	The paper undertakes a **case study** to determine the drivers, as well as the costs and benefits of incurring TD in the Change Control Board's decision process. The study is conducted in two stages: first focused on defect history records review for selected subject systems. A set of cost categories related to defect fixing, and their incurring conditions were identified. In the second stage 7 CCB members were interviewed, additional cost categories, and factors affecting the decisions about when to fix a defect were identified.	27
The Benefits and Consequences of Workarounds in Software Development Projects	2015	Yli-Huumo, Maglyas, Smolander [40]	The paper presents an **interpretive case study** on drivers, benefits, and costs of undertaking workaround solutions in software development projects. Semi-structured interviews are conducted with employees in two software development organisations. Interview results are grouped per workaround driver, with respective benefits and costs identified.	4

follow-up interview with the team leader afterwards revealed that the decision to implement more automated tests had to do with the difficulty (the "pain") of testing, but had more to do with improving the quality of testing." This could also suggest that a different form of reasoning was in place that may not weigh multiple criteria and evaluate multiple options to choose the options with the highest utility. Numerical optimization across multiple objectives is typically depicted as automatically superior to individual expertise and team knowledge, as in the following quote: "Technical debt…is currently managed in an implicit way, if at all. Decisions are largely based on a manager's experience, or even gut feeling, rather than hard data gathered through proper measurement." [12, :160] Guo's work brings out the tension between prescriptive and descriptive views when it discusses how "the actual decision making process, as observed in the case study, worked differently", in contrast to the assumptions and intentions of the method design. The method design was based on an approach that assumed the applicability of MCDA and computed a ranking of options ("we had designed our proposed approach so that all factors could be incorporated into the notions of principal and interest, and so principal and interest could then be used as the sole criteria in decision making") while the team "found it preferable to simplify the estimation of principal and interest (concentrating strictly on effort), but to be more holistic in actual decision making, by taking other factors into consideration at that time." Decision makers thus preferred other approaches for dealing with complexity. A similar experience arises in Oliveira [27]: The authors apply the framework proposed by Seaman and Guo, and the companies studied agreed on using it, but in the actual decision made about paying back debt, the decision making deviates from the method. The paper reports on the study of two teams. Neither of the teams prioritized technical debt as required. Even in situations when according to the proposed normative framework, the benefits were higher than the costs, the normative suggestion to allocate time to pay back debt was not adopted. Teams chose other options. For example, one team decided to allocate a fixed amount of hours to pay back debt.

Astoundingly, despite persistent deviations from the proposed norms, the use of normative assumptions is never questioned. Instead, deviations are treated as deficiencies that should be overcome. For example, Guo et al. [11] makes this explicit: Effectiveness and efficiency are taken as the only criteria for decisions. Other possible obstacles to method adoption, including culture, politics, group dynamics and individual resistance are not taken into account. The groups' deviation from the proposed model is instead written off with the statement "there was not enough time" [11, :164].

4.3 RQ3: The role of time in trade-off decisions

Table 2 summarizes the varied notions of time that surface across the analyzed 9 studies, which confirm and expand those previously observed in [5].

As expected, technical debt decisions frequently involve direct attention to the intertemporal and path-dependent nature of development. The trade-off between short-term needs and longer-term needs is explicit in several of the decisions that are examined in the studies. Examples of intertemporal choices studied in these papers include prioritization, release planning, and decisions to postpone the repaying of technical debt. In doing so, it is understood that decision makers have difficulties in estimating and valuing uncertain consequences occurring at different points in the future, but explicit attention to how these concepts of intertemporal choice can be expressed is generally absent. Yet, the studies provide interesting and insightful observations and findings.

Yli's study paints a rich picture of the role of time in technical debt decisions, using the concept of the 'workaround' [40]. An interesting aspect highlights that the valuation of time as a resource is not straightforward. More specifically, the value of time *over time* is not linear. For example, one studied company decided to implement a workaround to finish a previously announced feature in time, only to immediately invest a much larger amount of resources to fix the workaround right away upon the feature release. The reason for this was external to the project or the team, located in the business context. Yet, this example shows that the same *amount of time* has different value *at different points* in time, depending on other factors, which are highly contextual and not represented by the standard economic approaches based on linear discount factors [8]. This is consistent with a meta-review of intertemporal choice studies in economics that found a 'spectacular disagreement' in the future discount factors inferred from the actual choices people made in numerous empirical studies [10]. Yli's study highlights that in TD management, decisions are taken differently by different roles, that organizational roles and hierarchy and power are important for the outcome and should be studied, and that business interests often dominate over technological concerns, even if the business perspective is shorter- rather than longer-term.

Klinger's study highlights time as a factor of complexity, noting that "an assessment is complicated by the fact that technical debt depends heavily on many dynamic and challenging factors that may change over time, including customer requirements, dependencies between products and teams, ecosystem changes, and mergers and acquisitions". Furthermore, temporal relations can exist across systems, leading to "cascaded impact from decisions made on other projects on which a given project depends. This cascading effect may happen along interfaces between development groups or even temporally across the ecosystems that come to depend on the decisions from one release to another." Similar notions of effects accumulating in time are highlighted by Leppänen [20]: "When refactoring is overlooked in daily development, the likelihood of larger refactorings increases with time."

Despite the varied nature of time in these discussions, we note that the explicit conceptual frameworks of intertemporal choice research have not been taken into account by these studies. This suggests that future empirical studies of technical debt should incorporate research questions into their study designs that examine how the future is discounted by decision makers and other stakeholders [10] and more generally, what conceptualizations appear in the decision making process [24] and what role such factors as distance and affect [38] or framing [34] play in the priorities and preferences of decision making.

4.4 Threats to Validity

4.4.1 Construct Validity. **Technical Debt** is a very well established and consistently used term in software engineering. Correspondingly, we found few irrelevant papers included in this search.

The judgment of whether a given work is **empirical research** was somewhat more subjective at times. The researchers paid particular attention to this coding step to try and ensure that all relevant studies were captured. This led to the inclusion of three studies in the final stage of analysis that turned out not to investigate specific decisions empirically. These were excluded at that stage.

Coding of the underlying decision making theories and assumptions is arguably the most subtle and difficult of judgements, and hence was performed only on the small set of empirical studies– in research as in development, time is a limited resource. Independent coding throughout, detailed qualitative annotations, and discussions between all authors were performed to ensure consistency

Table 2: Varied notions of Time surface in the empirical studies.

Time as.../ in	Limited Resource	Time to Apply	Sequence of Project Events	Time to Market	Axis of Change	Other: Time as...
Bellomo [6]	X		X	X		The time of a deployment cycle, ideally to be shortened to shorten the feedback cycle; binding time (to be deferred sometimes); application runtime in distributed systems, with focus on issues of synchronization
Guo [11]	X	X	X			
Guo [12]	X	X				Project deadlines, including very hard deadlines that must be met at the risk of losing business. Current vs. future technical debt.
Klinger [18]	X		X			A factor of complexity - context factors change over time, and temporal relations can exist across systems
Leppänen [20]	X		X		X	An axis on which effects can accumulate
Oliveria [27]	X	X	X			Events recurring in time
Siebra [30]	X	X	X	X		The timeframe of the system development lifecycle; the time to complete before a deadline
Snipes [31]	X		X		X	Time as a factor of change; the object of estimation (estimate the time); the timeline of the research study
Yli [40]	X	X		X	X	The history of a discipline and its evolution; the value of time over time

and construct validity. To enable critique, we have provided ample quotes from the specific papers in the detailed discussion to highlight the basis upon which we have classified previous studies. Additionally, we did conduct cross validation and an initial test on three papers coded by all three researchers to further enhance our understanding and agreement.

Researcher expectations bias was considered, as one of the authors had conducted a previous study on the notion of time in SE, and the previously developed codeset was used as a starting point of the current analysis. To counter this threat, the two additional authors collaborated in this research with no previous work or expectations on the research outcomes. The preliminary codeset was not considered a limiting factor, but was expanded and revised as analysis progressed (e.g., adding such categories as time as investment, and time as risk factor).

4.4.2 Internal Validity. Although we followed Kitchenham's guidelines for systematic literature reviews, minor deviations from the protocol should be noted:

- We did not evaluate in detail the quality of the empirical work we studied. This was due to the fact that we concentrate on answering the RQs in our mapping study instead. In this process, however, we encountered papers describing their research as "case study research" that did not study decision making empirically. We marked these papers accordingly.
- We did not evaluate our review protocol using an external reviewer prior to analysis. We piloted our coding, however, and the protocol was developed in mutual agreement. Moreover, the initial search and the final protocol have been externally reviewed.

4.4.3 External Validity. The searches were limited to 3 databases, and no snowballing was conducted. This limits the external validity of our findings. However, the databases we used are commonly considered the main sources, and Google Scholar is often seen as 'most comprehensive' source.

The concept of technical debt is well established in the field and very consistently used, so that we are quite confident to have captured a representative body of literature. However, the conclusions drawn here may not generalize to broader studies of decision making in software engineering.

The search for data extraction was performed in December 2016. Very recent work will be missed from this report, which could threaten the up-to-date generalizability of the work. Given our study's focus on trends, however, this threat is limited. If a newer study takes naturalistic perspectives, in fact, it is in agreement with our recommendations.

4.4.4 Reliability. Three coders worked independently with regular check-ins. We performed the initial mapping in a total of eight rounds. In each, a set of papers was coded by more than one coder, followed by a discussion of discrepancy. We noticed a successive reduction of inconsistencies in the later rounds of coding.

For the detailed qualitative analysis of empirical papers, all papers were coded by at least two authors. Independent detailed qualitative annotations were produced, and discussions between all authors took place to ensure consistency in the application of concepts and coding. An arms-length review of the protocol was performed following the completion of a draft of the report.

5 RELATED WORK

The present SLR is closely related to our previous work that investigates how trade-off decisions (over time) have been conceptualized within SE literature [5]. When working on [5], an auxiliary search on 'technical debt' term was conducted as the concept of Technical Debt (TD) is closely related to the dimensions of trade-off decisions over time. TD includes a temporal dimension, manifesting itself in the concept of debt. Therefore, literature on TD was considered by the authors of [5] to be complementary to the literature found on trade-off decisions. However, the corpus on technical debt itself was not analyzed as part of [5], but only the overlap between the papers on technical debt and on trade-off decisions was noted. In [5] five overlapping papers were reported, out of which 2 have been classified as relevant for the present study.

In respect with findings, the current paper confirms results observed in [5] on the scarcity of empirical studies on time aspects in trade-off decision making in SE, as well as on the the dominance of normative decision models.

Value-based SE (VBSE) has suggested to integrate the values pertaining to software system into the engineering process, yet, predominantly the values are expressed in terms of utility functions [8]. In some VBSE work, the decision making process is viewed as naturalistic one, whereby the possible images of future are considered and a path to attaining the preferred ones is outlined [7], yet still, it is promoting the rational agent's model of decision makers.

The vast majority of TD literature, which is explicitly focused on considering decisions for making trade-offs across time [26], focuses on building normative models of the principal and interest calculation for this "debt". Accordingly, recent work is inspired by the automated trading environments of the current financial markets. We observe, however, that while it may make sense to assume that auto-trading at speculative markets can always algorithmically identify and select the current "best" decision as the one with the highest expected return, in contrast, TD decisions are taken in purposeful social systems within an evolving context of project circumstances. They are taken by humans rooted in their personal experiences and organizational situations.

6 CONCLUSIONS AND OUTLOOK

6.1 Summary

The above discussion shows that while some research on technical debt is empirical in nature, very few studies have been conducted that explicitly focus on examining decision making processes that incur or address technical debt in empirical research. Instead, most research is focused on developing methods, models, and metrics that are designed to feed into decision making approaches, without examining how exactly these decisions are made. Those studies that examine decisions in detail suggest that empirical research can provide deep and unexpected insights of profound relevance to the development of methods, models and metrics. However, a broader perspective on decision making theories seems necessary to take advantage of these empirical insights and situate them in decision making theory. From the mapping discussed above, the following key findings stand out:

(1) **Intertemporal choices are common in technical debt**. However, the nature of time has not been explored fully thus far, and the theory of intertemporal choice has not been brought to bear on decision making in software engineering - a finding that confirms our previous study [4]. The literature reviewed above characterizes intertemporal choices in technical debt from a variety of perspectives, and has brought forward unique insights and observations on the nature of decision making in Technical Debt management. Time plays a complex role in the empirical studies. However, the literature on intertemporal choice has not been considered in Technical Debt research. Considering that technical debt is in its essence an effect of intertemporal choices, this suggests a significant opportunity for future researchers.

(2) **Empirical research is scarce**: Most research in technical debt does not study the actual decision making processes empirically, but draws its assumptions from normative theories of decision making. The mapping discussed above and summarized here in Figure 3 shows that while some research on technical debt is empirical in nature, very few studies have been conducted that explicitly focus on examining decision making processes that incur or address technical debt in empirical research. Instead, most research is focused on developing methods, models, and metrics that are designed to feed into decision making approaches, without examining how exactly these decisions are made. More empirical research is thus needed to understand how various stakeholders in software projects take trade-off decisions about issues related to technical debt.

(3) **Broader theoretical perspectives are needed.** Those studies that examine decisions in detail suggest that empirical research can provide deep and unexpected insights of profound relevance to the development of methods, models and metrics. However, a broader perspective on decision making theories is necessary to take advantage of these empirical insights and situate them in decision making theory. The review above also significant opportunities to broaden the theoretical and methodological frame of TD research to incorporate perspectives beyond the dominant paradigm of rationalistic decision making theory, including naturalistic perspectives of descriptive decision making theory that are better able to express and explain some types of decision making processes as they occur in practice. This corresponds to recommendations made by Zannier et al. [41] in a study of decision making in software engineering.

6.2 Recommendations for Future Research

We suggest that the following priorities should be pursued in more depth to address the identified gap in empirical research and incorporate other disciplines' insights:

(1) Explicitly study decision making processes in their real context using observations and ethnographic approaches [29] to capture rich descriptions of real cases.
(2) In studying decision making empirically, refrain from predicating research questions and data collection purely on rationalistic approaches and instead, incorporate a broader perspective including descriptive decision making theories.
(3) Explicitly investigate how decision makers discount future outcomes in TD management.
(4) Examine how psychological distance affects the preferences of decision makers in TD management.
(5) Conduct experiments that examine how the framing of TD management decisions as losses or gains affects judgment and decision making in TD.

(6) Triangulate observations of decision making with post-hoc interviews, because decision maker's reconstructed memory of the decision making process often differs significantly from observations.

ACKNOWLEDGMENTS

The authors thank B. Penzenstadler, N. Seyff, and C. Venters for commenting on the data extraction and analysis protocols, and Dawn Walker for her support in data collection and preparation. This research has been partially supported by NSERC through RGPIN-2016-06640, by the Connaught Fund, and by the UK EPSRC Refactoring Energy Systems fellowship (EP/R007373/1).

REFERENCES

[1] Paris Avgeriou, Philippe Kruchten, Ipek Ozkaya, and Carolyn Seaman. 2016. Managing Technical Debt in Software Engineering (Dagstuhl Seminar 16162). *Dagstuhl Reports* 6, 4 (2016), 110–138. https://doi.org/10.4230/DagRep.6.4.110
[2] Lee Roy Beach and Raanan Lipshitz. 1993. Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. (1993). http://psycnet.apa.org/psycinfo/1993-97634-002
[3] Christoph Becker, Ruzanna Chitchyan, Stefanie Betz, and Curtis McCord. 2018. Trade-off Decisions Across Time in Technical Debt Management: Literature review Coding and Reference Documentation. (2018). https://doi.org/10.5683/SP/DGOIZ8
[4] Christoph Becker, Dawn Walker, and Curtis McCord. 2017. Intertemporal choice: decision making and time in software engineering. In *Proceedings of the 10th International Workshop on Cooperative and Human Aspects of Software Engineering*. IEEE Press, 23–29.
[5] Christoph Becker, Dawn Walker, and Curtis McCord. 2017. A systematic literature review on intertemporal choice in software engineering-protocol and results. *arXiv preprint arXiv:1701.08310* (2017).
[6] Stephany Bellomo, Neil Ernst, Robert Nord, and Rick Kazman. 2014. Toward Design Decisions to Enable Deployability: Empirical Study of Three Projects Reaching for the Continuous Delivery Holy Grail. In *Proceedings of the 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN '14)*. IEEE Computer Society, Washington, DC, USA, 702–707. https://doi.org/10.1109/DSN.2014.104
[7] Michael Berry and Aybüke Aurum. 2006. Measurement and Decision Making. In *Value-Based Software Engineering*. Springer, 155–177.
[8] Stefan Biffl, Aybuke Aurum, Barry Boehm, Hakan Erdogmus, and Paul Grünbacher. 2006. *Value-based software engineering*. Springer Science & Business Media.
[9] Neil A Ernst, Stephany Bellomo, Ipek Ozkaya, Robert L Nord, and Ian Gorton. 2015. Measure it? manage it? ignore it? software practitioners and technical debt. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*. ACM, 50–60.
[10] Shane Frederick, George Loewenstein, and Ted O'donoghue. 2002. Time discounting and time preference: A critical review. *Journal of economic literature* 40, 2 (2002), 351–401.
[11] Yuepu Guo, Carolyn Seaman, and Fabio QB da Silva. 2016. Costs and obstacles encountered in technical debt management–A case study. *Journal of Systems and Software* 120 (2016), 156–169.
[12] Yuepu Guo, Rodrigo Oliveira Spínola, and Carolyn Seaman. 2016. Exploring the costs of technical debt management–a case study. *Empirical Software Engineering* 21, 1 (2016), 159–182.
[13] Ralph L Keeney and Howard Raiffa. 1993. *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press.
[14] Barbara Kitchenham. 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University* 33, 2004 (2004), 1–26.
[15] Gary Klein. 2008. Naturalistic Decision Making. *Human Factors* 50, 3 (June 2008), 456–460. https://doi.org/10.1518/001872008X288385
[16] Gary A. Klein (Ed.). 1993. *Decision making in action : models and methods*. Ablex Pub., Norwood, N.J.
[17] Gary A Klein. 1999. *Sources of power: How people make decisions*. MIT press.
[18] Tim Klinger, Peri Tarr, Patrick Wagstrom, and Clay Williams. 2011. An enterprise perspective on technical debt. In *Proceedings of the 2nd Workshop on managing technical debt*. ACM, 35–38.
[19] Per Lenberg, Robert Feldt, and Lars Göran Wallgren. 2015. Behavioral software engineering: A definition and systematic literature review. *Journal of Systems and Software* 107 (2015), 15–37.
[20] Marko Leppänen, Samuel Lahtinen, Kati Kuusinen, Simo Mäkinen, Tomi Männistö, Juha Itkonen, Jesse Yli-Huumo, and Timo Lehtonen. 2015. Decision-making framework for refactoring. In *Managing Technical Debt (MTD), 2015 IEEE 7th International Workshop on*. IEEE, 61–68.
[21] Raanan Lipshitz, Gary Klein, Judith Orasanu, and Eduardo Salas. 2001. Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making* 14, 5 (Dec. 2001), 331–352. https://doi.org/10.1002/bdm.381
[22] George Loewenstein, Ted O'Donoghue, and Matthew Rabin. 2003. Projection Bias in Predicting Future Utility. *The Quarterly Journal of Economics* 118, 4 (2003), 1209–1248. https://doi.org/10.1162/003355303322552784
[23] George Loewenstein, Daniel Read, and Roy F Baumeister. 2003. *Time and decision: Economic and psychological perspectives of intertemporal choice*. Russell Sage Foundation.
[24] George Loewenstein, Scott Rick, and Jonathan D Cohen. 2008. Neuroeconomics. *Annu. Rev. Psychol.* 59 (2008), 647–672.
[25] Philipp Mayring. 2000. Qualitative Content Analysis. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, Vol. 1.
[26] Steve McConnell. 2007. Technical debt. *Software Best Practices, Nov* (2007).
[27] Frederico Oliveira, Alfredo Goldman, and Viviane Santos. 2015. Managing technical debt in software projects using scrum: An action research. In *Agile Conference (AGILE), 2015*. IEEE, 50–59.
[28] Johnny Saldaña. 2012. *The Coding Manual for Qualitative Researchers*. Number 14. Sage.
[29] Helen Sharp, Yvonne Dittrich, and Cleidson RB de Souza. 2016. The role of ethnographic studies in empirical software engineering. *IEEE Transactions on Software Engineering* 42, 8 (2016), 786–804.
[30] Clauirton Siebra A Siebra, Graziela S Tonin, Fabio QB Silva, Rebeka G Oliveira, Antonio LOC Junior, Regina CG Miranda, and Andre LM Santos. 2012. Managing technical debt in practice: an industrial report. In *Proceedings of the ACM-IEEE international symposium on Empirical software engineering and measurement*. ACM, 247–250.
[31] Will Snipes, Brian Robinson, Yuepu Guo, and Carolyn Seaman. 2012. Defining the decision factors for managing defects: a technical debt perspective. In *Managing Technical Debt (MTD), 2012 Third International Workshop on*. IEEE, 54–60.
[32] Evangelos Triantaphyllou. 2000. *Multi-Criteria Decision Making Methods*. Springer US, Boston, MA, 5–21.
[33] A Tversky and D Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211, 4481 (1981), 453–458. https://doi.org/10.1126/science.7455683
[34] Amos Tversky and Daniel Kahneman. 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review* 90, 4 (1983), 293.
[35] Amos Tversky and Daniel Kahneman. 1986. Rational Choice and the Framing of Decisions. *The Journal of Business* 59, 4 (1986), S251–S278.
[36] John Von Neumann and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton university press.
[37] John Von Neumann and Oskar Morgenstern. 1944. *Theory of games and economic behavior*. Princeton university press.
[38] Elke U Weber. 2006. Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). *Climatic change* 77, 1 (2006), 103–120.
[39] Roel Wieringa, Neil Maiden, Nancy Mead, and Colette Rolland. 2006. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. *Requirements Engineering* 11, 1 (2006), 102–107.
[40] Jesse Yli-Huumo, Andrey Maglyas, and Kari Smolander. 2015. The benefits and consequences of workarounds in software development projects. In *International Conference of Software Business*. Springer, 1–16.
[41] Carmen Zannier, Mike Chiasson, and Frank Maurer. 2007. A model of design decision making based on empirical results of interviews with software designers. *Information and Software Technology* 49, 6 (2007), 637–653.