Digital Nudging for Technical Debt Management: Insights from a Technology-driven Organization

Lorena Buchmann
University of St. Gallen
lorena.b@hotmail.com

Kazem Haki
University of St. Gallen
kazem.haki@unisg.ch

Abstract

Today's fast-growing and ever-changing business environments force software development teams to release high-quality software on a tight schedule. The notion of technical debt (TD) captures the technical compromises wherein software quality is sacrificed for short-term goals. One of the most significant challenges for technical debt management (TDM) is time as a factor of complexity. TD arises from decisions that are favorable in the short term but cause a need for complex and costly actions in the long term. Building on the applications of nudges – the use of small design modifications in choice environments to guide people's behavior – and on their psychological effects, in this design science research we develop and evaluate design elements of a Technical Debt Management Label (as a TDM nudge) in a technology-driven organization. The TDM label aims to guide software developers' decisions towards adopting TDM activities that are favorable in the long term.

1. Introduction

Today's fast-growing and ever-changing business environments force software development teams to release high-quality software on a tight schedule [1]. These time constraints create constant pressure on software development teams to quickly complete all their software development tasks [1]. Such pressure as well as increasing software development complexity often create a context in which the development team may choose to take shortcuts and workarounds and therefore assume some technical debt (TD) to bring about business value [2]. TD refers to problems faced during software development and maintenance considering tasks that are carried out inadequately and consequently cause higher costs and severe quality issues in the long run [3]. Development teams may lack awareness on the consequences or even existence of TD and thus, accrue TD unintentionally. Uncontrolled TD or lack of its awareness can have a significant, negative impact on information systems (IS) such as impaired team velocity, severe quality issues, inability to add new features, premature loss of a system, and increasing evolution costs [1]. TD management (TDM) decisions often involve a trade-off between uncertain future outcomes [4]. Regardless of the factors that motivate such decisions, such as tight schedules, TDM decision-making involves a trade-off in time, a "now vs. later" logic [4]. Examples of intertemporal choices include prioritization, release planning, and decisions incurring TD, such as using an old platform version when a new version is already available. For decision-makers, estimating and valuing uncertain consequences occurring in different points in time pose difficulties because these choices are complex and infrequent.

Recently, a promising approach to behavior change based on relatively cheap and easy alterations of features in decision contexts has sparked public interest. Thaler and Sunstein [5] posited the idea that human behavior is guided by the decision-making context. Described as "choice architecture" and "nudging," the basic concept underlying nudging is to leverage psychological effects from behavioral sciences to design decision-making contexts. Whereas the choice architecture is the context in which individuals make decisions, nudges are interventions within the choice architecture aimed at steering individual's behavior in specific directions [6].

TD research has largely developed sophisticated approaches for making TDM decisions, such as financial modeling techniques [4]. In this study, we posit that nudging offers a promising complementary approach to existing traditional decision-making models for TDM. Inspired by the framework for energy labeling of the European Union and the Enterprise Architecture Label by Schilling et al. [7], this study aims to design and implement design elements for a *Technical Debt Management Label*. Accordingly, we address the following research

URI: https://hdl.handle.net/10125/71115
978-0-9981331-4-0

question (RQ): *Which design elements should be included in a nudge for technical debt management in order to support TDM decisions among developers?* To address this question, we employ a design science research (DSR) approach [8] to develop and evaluate design elements for a TDM Label. The artifact is validated quantitatively through survey research and qualitatively through semi-structured expert interviews and focus group workshops in a technology-driven organization. The design of the artifact is guided by psychological effects underlying nudging and TDM research. This study offers both a demonstrated and evaluated digital nudging artifact as well as an initial set of design principles for the development of labels as digital nudge.

2. Research background

2.1 Technical debt management

During software development, development teams are faced with various challenges such as short deadlines and high-quality expectations [3]. In this context, tasks are sometimes carried out inadequately and some decisions are made with short-term benefit (e.g., shortened time to market), but requiring long-term maintenance effort later on in the software development cycle [3]. This phenomenon is known as TD. The TD metaphor was first introduced by Ward Cunningham in 1992 in his experience report on the delicate balance between code quality and short-term benefits: *"Shipping first-time code is like going into debt. A little debt speeds development so long as it is paid back promptly with a rewrite […] The danger occurs when the debt is not repaid. Every minute spent on not-quite-right code counts as interest on that debt"* [9, p. 30]. The metaphor involves an analogy to financial debt. However, the debt is not borrowed money but borrowed time, or specifically extra effort that can be translated into monetary terms [10].

Avoiding TD in a software system is impossible [11]. All organizations with software-intensive systems must live with different TD types (e.g., code debt, architectural debt, design debt) within their context [10]. Further, TD is not necessarily 'bad' for software development projects. To achieve short-term benefits, such as increasing software development velocity in order to deliver new features more rapidly, some intentional TD might be even fruitful [2]. Nonetheless, long-term and cumulative effects of unnecessary TD can lead to hardly controllable and detrimental consequences [12]. The latter manifests in reduced productivity and system fragility, and can even affect the development team's morale and motivation [12]. To this end, TDM is concerned with activities to manage, measure, prevent and reduce TD during software development [1]. Li et al. [2] identified eight TDM activities: TD identification, TD measurement, TD prioritization, TD prevention, TD monitoring, TD repayment, TD representation/documentation, and TD communication. For each identified TDM activity, Li et al. [2] have collected and synthesized different approaches that were developed, proposed, and used for TDM in the literature. Our study addresses the call by Li et al. [2] to develop more sophisticated TDM tools in order to manage different types of TD that can be integrated into the development teams' daily work environment [2]. To this end, this study develops and evaluates an artifact (a TDM label) based on TDM and nudging literature. The next section elaborates on the foundations of nudging.

2.2 Nudging

Behavioral sciences have revealed ample evidence that human behavior and choice is influenced by emotions, the use of mental shortcuts (heuristics), biases, and the decision-making context [13]. Thaler and Sunstein [5] coined the term "choice architecture" to reflect the fact that choices can be presented in different ways as well as the impact of how the choice is presented on the decision-maker. The concept of nudging states that subtle changes in the choice architecture can influence behavior. Its primary goal is to guide individuals towards desired choices and behaviors *without forbidding any options or significantly modifying the incentives* [5]. These changes in behavior and choice are guided by an understanding of psychological biases and heuristics. Nudging attempts to overcome or make use of these heuristics and cognitive biases to guide people's behavior in a predictable way [5, 14]. Table 1 presents some of the most frequently mentioned psychological effects underlying nudging.

Nudging has been shown to be appropriate in situations in which decisions are complex and its consequences are separated in time, choices are infrequent, as well as when feedback is slow or infrequent [5, 15]. Such characteristics of nudging fit well to TDM decisions as they always imply a trade-off between decisions that favor short-term outcomes and those with long-term disadvantages. While extant TDM models and approaches predominantly promote formally enforced procedures, in this study we offer nudging as an informal means (as they do not forbid any options) to complement existing approaches.

Table 1. Psychological effects underlying nudging

Framing	The framing effect states that decisions depend in part on the way choices are presented and phrased [16].
Social norms	Social norms are "rules and standards that are understood by members of a group and that guide and/or constrain social behavior without the force of laws" [17, p. 152].
Messenger effect	The messenger effect describes that information is weighed depending on the perceived authority of the source of that information [18].
Anchoring	Anchoring is a psychological effect where an individual makes estimates by using an initial starting point (anchor) [19].
Availability heuristics	Availability heuristics states that individuals' estimates about likelihood of events is influenced based on the "the ease with which instances or occurrences can be brought to mind" [19, p. 1127].
Loss aversion	Loss aversion describes the tendency of individuals to dislike losses more than gains of the same size [20].

3. Design science research approach

The main goal of this study is to design a digital nudge and evaluate which design elements should be included so that software development teams understand and become more aware of the TD in their system and consequently can take actions to manage it. We followed the DSR methodology proposed by Peffers et al. [8] to structure our research process as explained in the following.

Our research started with (i) defining the specific research problem and justifying the value of the solution. TD research proposes several formal approaches on how software developers and architects should make TDM-related decisions, but these proposals revolve around sophisticated financial modeling techniques to provide measurable statistics and indicators [4]. Further, formal approaches in TDM require enforcement mechanisms that may not be followed and appreciated by development teams when developing software under tight schedules. As a complement, informal mechanisms to nudge development teams towards intended behavior can cater an affective TDM approach. The development of a nudge to steer software developers' behavior would provide developers with relevant information and support the complex nature of TDM decisions. Subsequently, based on the problem definition, (ii) we inferred and defined the objectives for the solution. The objective is to develop and evaluate the design elements for a TDM Label (as a digital nudge) in a technology-driven organization that supports software development teams in making TDM-related decisions. In the next step, we (iii) developed the artifact and its design elements building on TD and nudging research. In the fourth step, the developed artifact is (iv) demonstrated to experts as well as to development teams of the case organization.

According to guidelines for DSR proposed by Hevner et al. [21], evaluation is a crucial element of the DSR process. This study follows an observational method to demonstrate and evaluate the artifact and its effects in a real-world situation [21]. To this end, we selected a large and technology-driven organization[1], since it has a long tradition in software development and offered the need and conditions to design an artifact for TDM. Time-to-market pressure forces their developers to realize short-term benefits on the expense of long-term sustainability. Thus, the case organization started a project to develop a TDM approach in August 2019.

The (v) evaluation activity of the TDM Label involved three activities. Firstly, we evaluated the label qualitatively through semi-structured expert interviews at the case organization to collect insights and feedback. This evaluation activity is associated with assertion and formal proof in order to improve and validate the artifact [22]. For the interview process, we followed the guidelines of Brinkmann [23]. After a brief introduction to our research project, experts were asked to evaluate the design of the label based on three evaluation criteria: understandability (i.e., the degree to which the label is comprehended), simplicity (i.e., the degree to which the label contained the minimal number of elements), and effectiveness (i.e., the degree to which the label can be supportive in TDM decision-making) [7, 22]. Further, suggestions for improving the label were collected. In total, twelve expert interviews were conducted during the period from February 2020 to April 2020 with senior architects and senior IT managers of the case organization who have an extensive knowledge of TD. The conducted interviews typically lasted between 20 and 37 minutes. Notes were taken and subsequently analyzed.

Secondly, in order to evaluate how well the TDM Label was received and understood by the target users, i.e. development teams, three focus group workshops were conducted. This form of qualitative feedback is also useful for exploring and examining different points of views, beliefs, needs, concerns, and experiences [24]. To guide the focus group workshop, the guidelines on using focus groups in DSR proposed by Tremblay et al. [25] were followed. Two groups with eight participants and one group with nine

[1] In compliance with company policies, the selected case organization is strongly anonymized.

participants were recruited in the case organization. Before the focus groups were conducted, the interview guide, which sets the agenda for a focus group discussion, was developed [25]. A series of three focus group workshops was conducted between March and April 2020. According to Tremblay et al.'s [25] suggestion, the workshop began with an explanation of the motivation behind the design of the TDM Label, proceeded to descriptions of how the artifact can be utilized and details of design, and concluded with the evaluation of the label. During the focus group workshops, which each lasted around 60 minutes, notes were taken and subsequently analyzed.

Last, the artifact's applicability and usefulness were assessed through a pre-post-survey design with the software development teams of the case organization. The data was obtained two times from the same subjects (within-subject comparison). The first survey (pre-survey) was distributed *before* the demonstration of the TDM Label to the development teams of the focus group workshops. *After* one week of the demonstration of the label, the second survey (post-survey) was distributed to the same participants. The survey creation followed the survey research process by Kasunic [26]. The designed questionnaire comprises four sections. The first part of the pre-survey questionnaire includes questions about demographic characteristics regarding participants' experience in software development, their project role, the size of the team, and whether their team is geographically distributed. The subsequent part measures the constructs TD Awareness, TD Understanding, and TDM Action, informed by Schilling et al. [7], according to which the statements are categorized. The measurement items are derived from the TDM literature and included statements about TDM activities [2] consisting of TD identification, measurement, prioritization, monitoring, repayment, representation, and communication. The statements provide answer choices on a seven-point Likert scale (1 = strongly disagree to 7 = strongly agree). The survey contains a total of 17 statements. However, only in the pre-survey questionnaire participants are asked to fill out the demographic information as the same participants fill in the post-survey questionnaire. After testing the questionnaire thoroughly by six developers of the case organization, the pre-survey questionnaire was distributed. The questionnaire format was paper-based and emailed to participants. After the nudge intervention, the post-survey questionnaire was distributed to the same participants as for the pre-survey questionnaire. In total, data from 16 respondents (67% response rate) was collected.

After collecting the results from the demonstration and evaluation steps, we also derived a set of design principles based on the method of Chandra et al. [27] for the development of labels as digital nudge.

4. Results

4.1 Technical debt measurement system in the case organization

The case organization selected four TD types based on their organization-specific ability to manage and influence. Thus, an overall TD index has been defined for all of the organization's IS to monitor TD evolution. The overall index is calculated based on code quality debt, infrastructure lifecycle debt, vulnerability debt, and automation debt. To quantify, these four index values are calculated on a normalized scale ranging from 1 to 10 (1 being low debt and 10 being high debt):

- Code quality debt: Refers to issues such as code smells, defect bugs, and security vulnerabilities.
- Infrastructure lifecycle debt: Includes after End-of-Service-Life (EoSL) date of application components (e.g., hardware) and termination of support by the component vendor. TD starts to increase because upgrades become more difficult when multiple versions need to be bypassed.
- Vulnerability debt: Application vulnerabilities cause a security risk to the case organization. Most vulnerabilities at the case organization can be fixed by upgrading components.
- Automation debt: Automating software development processes such as build and deployment processes require investment. Postponing these automation processes increases costs for the case organization due to required manual effort for build and deployment.

4.2 Development of the Technical Debt Management Label

Figure 1 depicts the developed artifact for an exemplary application A1. Thanks to its digital nature, the depicted values on the artifact are automatically fed by real-time data on the ever-changing TD measures to provide the real-time status of the measured application. The TDM Label was prototyped with *Sketch* and implemented in *Tableau* in the case organization. The design elements of the label are driven by psychological effects underlying nudging and TDM activities. What follows is an overview of the design elements as well as the theoretical reasoning that governed the design of the elements.

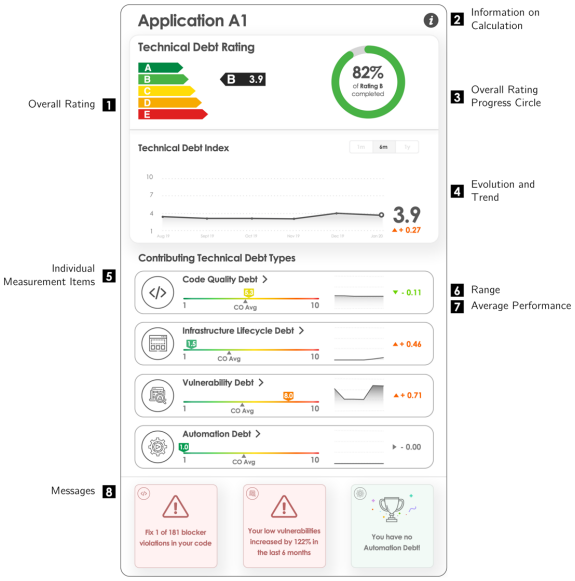

Figure 1. Technical Debt Management Label

Overall rating (DE01): At the top, the label provides the overall rating for TD for the respective application, communicated with a ranking and color code. The best TD-rated applications are in the green-colored ranking and the worst are in the red-colored ranking. This design feature is inspired by the successful framework for energy labeling of the European Union and the Enterprise Architecture Label [7]. An A to E scale is used for the overall rating of the TDM Label. Based on the overall TD index ranging from 1-10 for each application, the overall rating is calculated. For example, the best 15% of applications regarding TD index receive the rating A. This design element is governed by the psychological effect *framing*. The overall rating shifts the attention of developers by presenting the existing TD information in a different way. Further, by simplifying and framing complex information regarding the accumulated TD, it reduces the need for high cognitive capacity. Moreover, the overall rating design element is governed by the psychological effect *social norms*. The rating provokes the comparison to the other applications of the case organization that may informally force and encourage owners of low-ranked applications to take some actions and to improve their applications' overall rating.

Information on calculation and data sources (DE02): Developers can click the "*i*" in order to see more details on the calculation and data sources of the label. Such provision of details in the background ensures that the label as such is not crowded with information. This design element is based on the *messenger effect*. Generally, the weight given to information relies upon the perceived authority of the source of information [18]. Messenger credibility is particularly important: individuals are more likely to incorporate information into decision-making from sources deemed credible. Therefore, it is crucial to indicate the data source in the label in order to increase the probability of incorporating information of that label into TDM decisions.

Overall rating progress circle (DE03): The overall rating progress circle provides more detailed information on the completion of the overall rating (DE01). This design element is governed by the psychological effect *framing* and *loss aversion*. The progress circle is based on positive attribute framing. Positive as it refers to completion of the rating, i.e., "x% of Rating X completed." However, if the rating is very low, e.g., "14% of Rating X completed," the potential loss of the current rating could trigger loss aversion because, according to nudging theory, individuals have a tendency to dislike losses more than they like equivalent gains [28].

Evolution and trend (DE04): One of the TDM activities is to control the change of TD over time [2]. The evolution graph at the top of the label provides a graph showing the evolution of the overall TD index. The evolution graph is also displayed for each TD type that contributes to the overall TD. The time frame (i.e., 1 month, 6 months, 1 year) can be individually selected, which helps the developers visually monitor the evolution of TD in their application. Further, the trend is shown for the overall TD index as well as for every TD type depending on the selected time frame. This design element is governed by the psychological effect *loss aversion*. Most people feel that losses are more distressing than gains of equivalent size. Therefore, if the evolution of TD has an increasing trend, i.e., the accumulated TD increases, developers tend to be "loss averse" and overweigh potential losses. As a result, loss aversion causes developers to be biased towards choices that forgo such a loss aversion, e.g., taking decisions to reduce TD.

Individual measurement items (DE05): The individual measurement items design element refers to the four TD types that contribute to the overall application's TD index. This design element refers to the activity of TD measurement for quantifying the accumulated TD. Further, it allows one to monitor each TD type to verify if they are increasing or decreasing. Making types of TD visible can support decisions on repayment. However, TD types must be brought into a form that is useful for decision-making [29]. Therefore, the indexes are normalized on a 1-10 scale, providing comparability and facilitating communication at application level. The individual measurement items are governed by the psychological effect *framing*. The developer's choice architecture is influenced by the

different TD types that contribute to the overall TD index. Therefore, it guides the developers in their decision-making process regarding which TD type to focus on first. Existing information, i.e., the accumulated TD, is thus translated into a simplified form to reduce the cognitive effort.

Range (DE06): It is important to make TD visible because otherwise significant problems arise. Therefore, it needs to be ensured that TD achieves adequate visibility so that developers can consider it in decision-making processes [29]. Therefore, the metric is shown not merely as a number but also on a color-coded range from 1 to 10 (1 being low debt and 10 being high debt). This design element makes use of *anchoring*. The range of 1 to 10 as well as the position on the range serve as a source of reference and indicate the performance of each TD type.

Average performance of other teams (DE07): *Social norms* can have a potent influence on the individual's behavior [30]. Therefore, the average performance of other teams of an application for every TD type is shown. This element refers to descriptive social norms that "involve perceptions not of what others approve but of what others actually do" [31, p. 263]. The triangle shows the average performance of the other teams of the case organization for code quality debt, infrastructure lifecycle debt, vulnerability debt, and automation debt. Providing such a social reference point supports the developer's decision-making in order to judge the level of their own application.

Messages (DE08): Prioritization involves a crucial activity in TDM due to the limited resources for handling TD [12]. In general, alerts can be configured for the four measurement items when the metrics reach a certain threshold. In order to facilitate prioritization of TD, the label shows messages with TD focal points and/or achievements. The attention of developers can be triggered by using large fonts, bold letters, and bright colors [14]. The message "Your low vulnerabilities increased by 122% in the last 6 months," makes use of loss *framing* in order to nudge developers to fix the vulnerabilities. The message "181 blocker violations in your code" and "You have no Automation Debt!" makes use of *availability heuristics* that give more weight to readily available information.

4.3 Demonstration and evaluation of the Technical Debt Management Label

The initial demonstration and presentation of the TDM Label involved IS senior managers of the case organization. Taking place directly after each individual demonstration between 24.02.2020 and 15.04.2020, the evaluation aimed at assessing how well the TDM Label was received and understood by the experts.

The label was received very positively among the experts, who especially valued the simplicity and clearness of the label, as noted by one interviewee, *"I think it's clear, I think it's consumable. I like the visualization very much."* The results indicate that the quality of the visualization attracted predominantly positive acclaim. One point raised during the demonstration and evaluation concerned the relative nature of the overall rating (DE01). Despite the immediate association of the overall rating with the energy efficiency label, as noted by a participant, *"it's great, it's understandable, it's like an energy label"*, several participants raised the question of how the rating was calculated and what were the thresholds between the different ratings from A to E. The participants raised the concern that the relative rating could become contentious or a source of debate because it drives competition in the organization. One main point of feedback consisted of making the overall rating an absolute rating. As one participant commented, *"once you get the transparency, its impact on the behavior will become clear."*

The individual measurement items (DE05) were perceived very positively and no participant raised concerns about their understandability. Especially the messages (DE08) received positive acclaim among the experts and were deemed a powerful mechanism: *"I think there is power in this as a communication mechanism."*

Importantly, the label received positive affirmation as to its potential for guiding behavioral change in teams towards improving the rating, as noted by a participant: *"If you bring people to see their own rating, then obviously people want to do things to improve their rating."* Participants also highlighted that the label would be of great help in explaining TD to the business side. Thus, it can be concluded that the evaluation of the TDM Label by experts resulted in a positive outcome and the usefulness of the label was assessed as high.

The TDM Label has also been demonstrated in the context of focus group workshops with three different development teams within the case organization. Each development team was presented with the label with data for their application. Demonstration and subsequent evaluation with development teams took place on 03.03.2020, 30.03.2020, and 16.04.2020 and lasted 60 minutes per session.

Similar to the expert interviews, the resemblance of the TDM Label to the energy efficiency label received very positive affirmation among the development teams. Several participants pointed out that they liked this resemblance. Further, the visualization of the TDM

Label was deemed very high among the development teams. Several participants highlighted the simplicity and clearness of the label: *"The design is clear and transparent and very consistent because we have only a few metrics."* The results indicate that the label was very well received regarding simplicity and clearness. The developers viewed the evolution and trend (DE04) design element, with its adjustable time frame, as very useful for making TD-related decisions. Some points for consideration were raised regarding the understandability of the design elements. Similar to the expert demonstration and evaluation sessions, the overall rating (DE01) was subject of discussion during the demonstration and evaluation. An additional minor point concerned the average performance of other teams (DE07). Two participants raised the concern that it was not clear whether it was the average of the whole organization or the average of their team. Overall, the artifact was deemed very useful in supporting TD discussion: *"In general, I think that your dashboard shows information at the right level to stimulate or support a discussion with business."*

The execution of the TD survey was conducted at the case organization from 25.02.2020 to 01.05.2020 with the same development teams of the demonstration activity. Survey invitations were sent by e-mail to the software professionals working in agile software development teams. In total, the survey invitation was sent to 24 professionals. 22 of them completed the pre-survey questionnaire, and 16 completed the post-survey questionnaire, resulting in a response rate of 67%.

Most participants are Developers/Architects (75%), whereas 25% are assigned to the role IT Owner/Product Owner. Participants selected from among the following options to define their years of experience in their role: less than 2 years, 2 to 5 years, 6 to 10 years, or more than 10 years. Most have more than 10 years of experience in software development (56%), while some have six to ten years (19%), two to five years (19%), or less than two years of experience (6%) in software development. Clearly, the professionals who filled in the questionnaire have experience in software development. Concerning the size of the agile project teams, most of the participants work in a team consisting of 6 to 10 people (69%), though 31% work in a team consisting of one to five developers. Finally, all participants indicated that they work in geographically distributed teams.

The quantitative evaluation of the TDM Label shows an increase, on average, in all three constructs of TD Awareness, TD Understanding, and TDM Action *after* the demonstration of the label. Specifically, the label increased participants' awareness on the evolution over time of the TD in their application (56% agreement in the pre-survey, 63% agreement in the post-survey). Moreover, awareness increased for different types of TD (50% agreement in the pre-survey, 88% agreement in the post-survey). Furthermore, whereas before the nudge intervention, 0% of participants were aware of other teams that successfully manage TD, after the intervention, 38% of participants indicated that they are aware of other teams that successfully manage TD. Moreover, the label increased understanding of how to prioritize remediation of TD items (56% agreement in the pre-survey, 69% agreement in the post-survey). Before the nudge intervention, 25% of respondents indicated that their team monitors the evolution of TD over time. After the nudge intervention, all participants indicated willingness to monitor TD over time. Furthermore, before the intervention, 13% of participants indicated that they use visualization tools to communicate TD towards relevant stakeholders. After the intervention, 75% indicated willingness to use such tools to communicate TD issues. Based on 16 software professionals who participated in the survey before and after the nudge intervention, the quantitative results indicate the impact of the design elements of the label.

4.4 Reflection and learning

Collectively, the evaluation of the design elements resulted in a positive outcome. The visualization of the artifact in particular was assessed to be very high by both experts and development teams. The final version of the constructed artifact (see Figure 1), already incorporates feedback from the demonstration and evaluation phases.

Originally, the overall rating (DE01) was designed as a relative rating. By addressing the feedback from the expert interviews and focus group workshops, the overall rating (DE01) is now calibrated based on thresholds and is stable. Furthermore, the TD index (i.e., 3.9) of the application was inserted in the black arrow that indicates the rating category of the overall rating (DE01), clarifying how the evolution (DE04) relates to the overall rating (DE01). So as not to confuse users between the TD index with the overall rating from A to E (DE01), the title "Technical Debt Index" was inserted above the evolution (DE04) graph. Furthermore, the trends (DE04) were adapted to an absolute change instead of a change in percentage and have been placed on the right side of the label. Hyperlinks to the dashboards of the data source are indicated by an arrow next to each individual measurement item (DE05). Further, to indicate that the average performance of other teams (DE07) refers to the organizational average and not team average, the company abbreviation was added to "Avg." Feedback

regarding the messages (DE08) was incorporated by making the messages box smaller and making more actionable statements.

4.5 Design principles

In this section, the essential design knowledge that resulted from the evaluation activity is synthesized into design principles for the development of labels as digital nudge following Chandra et al.'s [27].

The design principles are based on the findings of the expert interviews and focus group workshops. A major need among experts and development teams is to have the possibility of intuitively drilling down to the specific dashboards of the data. This addresses the fact that the interviewees would like to have a clear action for each measurement item, allowing users to do the correct action to improve their measurement items. This requirement is reflected in **DP01 – interactivity:** *Provide a digital nudge that is interactive in order for users to intuitively drill down to the data and take action.* Moreover, the experts and developers demand timely and high-quality data. Otherwise the nudge loses its impact. This requirement is reflected in the design principle **DP02 – data quality:** *Provide a digital nudge that is connected with backup data with appropriate data quality in order for users to receive relevant insights at the right point in time.* Furthermore, a major need of the interviewees was to monitor the evolution of the measurement items in order to judge current and past performance and consequently take action, a point reflected in design principle **DP03 – evolution:** *Provide a digital nudge with an evolution graph in order for users to actively monitor their past and current development of metrics.* The fourth design principle reflects the need of the interviewees to know the exact calculation scheme of the individual measurement items: **DP04 – trust:** *Provide a digital nudge that has backup information of the details in order for users to receive a transparent and shared understanding of the design elements.* Thus, the backup information should be provided. The nudge should also provide the individual selection of time frames, which was mentioned multiple times in the interviews: **DP05 – adaptation:** *Provide a digital nudge that allows time frame selection in order for users to specify the time horizon needed for the analysis.* Doing so allows the users to adapt the time frame to their specific needs. Most of the experts and developers mentioned that they would like different granularity levels in order to allow for different views. This is reflected in **DP06 – perspective:** *Provide a digital nudge that allows the choice of different granularity levels in order for users to take different views.*

5. Discussion and conclusion

TDM decision-making remains a core challenge across software-intensive companies because it always implies the notion of time and trade-offs [4]. Several studies have introduced formal procedures to TDM decision-making [2, 4]. The concept of nudging [5] poses a complementing approach by guiding behavior based on easy alterations of features in decision environments. This study aims to develop design elements for a TDM Label, as a digital nudge, to support software developers in TDM decisions. The development and evaluation of design elements was structured according to the DSR process proposed by Peffers et al. [8]. This process was followed closely to ensure synergy between practical utility and rigor and yielded a precise description of design elements for a TDM Label along with an initial set of design principles for labels as digital nudge.

The qualitative results (expert interviews, focus group workshops) of the evaluation activity revealed that the effectiveness of the label was deemed very high by both IS experts as well as software developers of the case organization. The descriptive statistics of the quantitative evaluation activity (survey) performed during the DSR process indicate that the design and implementation of the TDM Label are appropriate for supporting TDM decisions and facilitate the assimilation process of TDM practices. Furthermore, survey results indicate that the label fosters awareness and understanding of TD topics among developers and suggests a high willingness to adopt TDM activities and communicate them to relevant stakeholders. Due to the complex nature and intertemporal component of TDM decisions [4], the label provides a decision tool that can change cognitive representation of intertemporal decisions to alleviate this bias [32].

This study contributes to the first design and development of a digital nudge for TDM, and a qualitative and quantitative evaluation of its elements in a technology-driven organization. The derived results in this study extend prior research on the design and use of digital nudges (e.g., [33]) by introducing a complementary approach that targets decision-making processes instead of using complex models to assist TDM decisions. Until now, software visualization techniques that support TDM have been studied in a very limited way [11]. This study moves the conversation forward by contributing to software visualization techniques in order to support TDM. A wide consensus exists that one of the major causes of TD is rooted in business [10]. Communicating and explaining TD to nontechnical stakeholders due to different knowledge backgrounds poses a major challenge in TDM. Shared artifacts can help to bridge

knowledge boundaries between the involved actors to enhance a common understanding [34, 35]. The TDM Label thus can be considered a boundary object to foster effective collaboration among the involved stakeholders [35].

Although this research was conducted with rigor and relevance, some limitations must be acknowledged. The major shortcoming of this study is its size restriction, with a sample of 16 developers used for the quantitative pre-post survey evaluation. The main reason for the relatively low number is the low response rate of development teams. The sample size limits the generalizability of the results based on the survey. Nevertheless, the selected survey participants were software professionals with extensive experience in IS. Another drawback of the study is the missing control group, the inclusion of which would show that the results can be attributed to the label. However, a control group was not feasible due to the small sample size. Furthermore, the underlying psychological effects for the label design were chosen based on the relevance to our context. Including all of them in one label was deemed unfeasible. Finally, given that this study is based on nudging theory, some ethical implications of nudging have to be considered. The main point of criticism regarding behavioral interventions is that it leverages automated and affective-cognitive processes to steer people's behavior unconsciously in a certain direction by exploiting biases [13]. Therefore, Sunstein [36] highlights the importance of transparency and awareness when designing behavioral interventions.

In light of the positive evaluation of the TDM Label and including the consideration of the discussed limitations elaborated above, future iterations of the DSR process can improve the label and push the discussion into new directions. The qualitative evaluation proved that the TDM Label as a whole received very positive acclaim and that the potential effectiveness of the nudge was high. Furthermore, the quantitative evaluation showed an increase in TD Awareness, TD Understanding, and TDM Action within the case organization. However, to reveal which of the design elements had the biggest impact on TDM decisions, one would need to analyze them individually. It was initially beyond the scope of this research to show the long-term effects of the label. Accordingly, as a next step, we need more research to establish whether the digital nudge is impactful in the long run regarding the overall rating (DE01) and the individual measurement items (DE05). This would allow us to analyze the quantitative long-term impact of the TDM Label as an intervention and analyze whether the TD index (i.e., the subject matter of the respective intervention) indeed decreases. Furthermore, we consider, and encourage prospective research, to implement the same intervention in other case companies to extend the data sample and the TD types.

Acknowledgements

This work has been supported by the Swiss National Science Foundation (SNSF).

References

[1] J. Yli-Huumo, A. Maglyas, and K. Smolander, "How do software development teams manage technical debt?–An empirical study", Journal of Systems and Software, vol. 120, 2016, pp. 195-218.

[2] Z. Li, P. Avgeriou, and P. Liang, "A systematic mapping study on technical debt and its management", Journal of Systems and Software, vol. 101, 2015, pp. 193-220.

[3] N. Rios, M. G. de Mendonça Neto, and R. O. Spínola, "A tertiary study on technical debt: Types, management strategies, research trends, and base information for practitioners", Information and Software Technology, vol. 102, 2018, pp. 117-145.

[4] C. Becker, F. Fagerholm, R. Mohanani, and A. Chatzigeorgiou, "Temporal discounting in technical debt: how do software practitioners discount the future?," presented at the 2019 IEEE/ACM International Conference on Technical Debt (TechDebt), 2019.

[5] R. H. Thaler and C. R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness, Penguin, 2009.

[6] A. Barton and T. Grüne-Yanoff, "From libertarian paternalism to nudging–and beyond", Review of Philosophy and psychology, vol. 6, no. 3, 2015, pp. 341-359.

[7] R. Schilling, S. Aier, and R. Winter, "Designing an Artifact for Informal Control in Enterprise Architecture Management," presented at the 40th International Conference on Information Systems (ICIS 2019), Munich, Germany, 2019.

[8] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research", Journal of Management Information Systems, vol. 24, no. 3, 2007, pp. 45-77.

[9] W. Cunningham, "The WyCash portfolio management system," in *Addendum to the proceedings on Object-oriented programming systems, languages, and applications (Addendum)*, Vancouver, British Columbia, Canada, 1992: Association for Computing Machinery, pp. 29–30.

[10] P. Kruchten, R. Nord, and I. Ozkaya, Managing Technical Debt: Reducing Friction in Software Development, Addison-Wesley Professional, 2019.

[11] T. S. Mendes, F. G. S. Gomes, D. P. Gonçalves, M. G. Mendonça, R. L. Novais, and R. O. Spínola, "VisminerTD: a tool for automatic identification and interactive monitoring of the evolution of technical debt items", Journal of the Brazilian Computer Society, vol. 25, no. 1, 2019, p. 2.

[12] E. Tom, A. Aurum, and R. Vidgen, "An exploration of technical debt", Journal of Systems and Software, vol. 86, no. 6, 2013, pp. 1498-1516.

[13] T.-B. Lembcke, N. Engelbrecht, A. B. Brendel, B. Herrenkind, and L. M. Kolbe, "Towards a Unified Understanding of Digital Nudging by Addressing its Analog Roots," in *Proceedings of Pacific Asia Conference on Information Systems*, 2019.

[14] C. R. Sunstein, "Nudging: a very short guide", Journal of Consumer Policy, vol. 37, no. 4, 2014, pp. 583-588.

[15] F. Ölander and J. Thøgersen, "Informing versus nudging in environmental policy", Journal of Consumer Policy, vol. 37, no. 3, 2014, pp. 341-356.

[16] A. Tversky and D. Kahneman, "The framing of decisions and the psychology of choice", Science, vol. 211, no. 4481, 1981, pp. 453-458.

[17] R. B. Cialdini and M. R. Trost, "Social influence: Social norms, conformity and compliance", in The handbook of social psychology, McGraw-Hill, 1998, pp. 151-192.

[18] P. Dolan, M. Hallsworth, D. Halpern, D. King, R. Metcalfe, and I. Vlaev, "Influencing behaviour: The mindspace way", Journal of Economic Psychology, vol. 33, no. 1, 2012, pp. 264-277.

[19] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases", Science, vol. 185, no. 4157, 1974, pp. 1124-1131.

[20] D. Kahneman, J. L. Knetsch, and R. H. Thaler, "Anomalies: The endowment effect, loss aversion, and status quo bias", Journal of Economic Perspectives, vol. 5, no. 1, 1991, pp. 193-206.

[21] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research", MIS Quarterly, 2004, pp. 75-105.

[22] C. Sonnenberg and J. Vom Brocke, "Evaluation patterns for design science research artefacts," in *European Design Science Symposium*, 2011, pp. 71-83.

[23] S. Brinkmann, Qualitative interviewing, Oxford University Press, 2013.

[24] P. Liamputtong, Focus group methodology: Principle and practice, Sage Publications, 2011.

[25] M. C. Tremblay, A. R. Hevner, and D. J. Berndt, "Focus Groups for Artifact Refinement and Evaluation in Design Research", Communications of the Association for Information Systems, vol. 26, 2010, pp. 599-618.

[26] M. Kasunic, Designing an effective survey, Pittsburgh, 2005.

[27] L. Chandra, S. Seidel, and S. Gregor, "Prescriptive knowledge in IS research: Conceptualizing design principles in terms of materiality, action, and boundary conditions," in *2015 48th Hawaii International Conference on System Sciences*, 2015, pp. 4039-4048.

[28] D. Kahneman, Thinking, fast and slow, Macmillan, 2011.

[29] N. Brown *et al.*, "Managing technical debt in software-reliant systems," in *Proceedings of the FSE/SDP workshop on Future of software engineering research*, 2010, pp. 47-52.

[30] R. B. Cialdini and N. J. Goldstein, "Social influence: Compliance and conformity", Annual Review of Psychology, vol. 55, 2004, pp. 591-621.

[31] R. B. Cialdini, "Descriptive Social Norms as Underappreciated Sources of Social Control", Psychometrika, vol. 72, no. 2, 2007, pp. 263-268.

[32] R. K. Ratner *et al.*, "How behavioral decision research can enhance consumer welfare: From freedom of choice to paternalistic intervention", Marketing Letters, vol. 19, no. 3-4, 2008, pp. 383-397.

[33] C. Schneider, M. Weinmann, and J. Vom Brocke, "Digital nudging: guiding online user choices through interface design", Communications of the ACM, vol. 61, no. 7, 2018, pp. 67-73.

[34] U. Gal, K. Lyytinen, and Y. Yoo, "The dynamics of IT boundary objects, information infrastructures, and organisational identities: the introduction of 3D modelling technologies into the architecture, engineering, and construction industry", European Journal of Information Systems, vol. 17, no. 3, 2008, pp. 290-304.

[35] K. Haki, S. Aier, and R. Winter, "A Stakeholder Perspective to Study Enterprisewide IS Initiatives," in *Proceedings of the 24th European Conference on Information Systems (ECIS 2016)*, Istanbul, Turkey, 2016.

[36] C. R. Sunstein, "Nudging and choice architecture: Ethical considerations", Yale Journal on Regulation, Forthcoming, 2015.