Phantoms of Innovation: Disciplined Simulation for Ex-ante Evaluation in Design Science Research

Philipp Matter
Zurich University of Applied Sciences, philipp.matter@zhaw.ch

Peter Heinrich
Zurich University of Applied Sciences, peter.heinrich@zhaw.ch

Katja Kurz
Zurich University of Applied Sciences, katja.kurz@zhaw.ch

Gianluca Miscione
Zurich University of Applied Sciences, gianluca.miscione@ucd.ie

Phantoms of Innovation: Disciplined Simulation for Ex-ante Evaluation in Design Science Research

Completed Research Paper

Philipp Matter, Peter Heinrich, Katja Kurz
School of Management and Law,
Zurich University of Applied Sciences
Theaterstrasse 17, 8401 Winterthur, Switzerland
{matp, heip, kurk}@zhaw.ch

Gianluca Miscione
University College Dublin &
School of Management and Law,
Zurich University of Applied Sciences
Belfield Campus, Dublin, Ireland
gianluca.miscione@ucd.ie

Abstract

For years, there has been an emphasis on how to efficiently and effectively identify, evaluate, and implement innovative information systems in both design science research (DSR) and practice. Nonetheless, still today, these efforts continue to be hampered by the temporal gap between ideation and evaluation. Usually, innovative ideas are implemented at a late stage of maturity (e.g., prototypes) to test their viability in practice. This widespread approach results in waste of resources and time if the viability of an idea fails outside the lab environment. This paper discusses an ex-ante evaluation approach derived from "pretotyping" that allows innovative ideas to be tested in naturalistic settings even before they have been implemented. Thus, we call them "phantoms". We show how this approach reduces temporal and relevance gaps, and we provide a preliminary assessment of its practicability by presenting and discussing three case studies conducted with real organizations and prospective users.

Keywords: pretotyping, ex-ante evaluations, innovation, simulation, design science

Introduction

The gap between design and use of information systems is central for digital innovation research and design science research (DSR). While digital innovation research leans toward a behavioral approach to understand *how technology is used and changes*, DSR takes a constructive approach on *how to reduce the gap between design and practice*. Compared to previous waves of technological innovation, mostly centered on influential R&D departments, digital innovation tends to originate from the fringes of open digital networks (Zittrain 2006). This means that countless innovations are tried out in a highly decentralized fashion. Studies show that up to 90% of innovation projects fail completely or partly (Rhaiem and Amara 2021, p. 190), leaving only a few to succeed. The consequences of this trial-and-error mode of innovation are several. Investors (especially venture capitalists) need to pay for their inevitable frequent losses with a few highly successful ventures. Accordingly, "fail fast" recommends shortening the cycle of trying things out in practice, and the introduction of agile methodologies has been inevitable for coping with the fast-moving design and development shift of goals – giving up many of the advantages of traditional (more sequential) system development processes. Despite this, gaps persist between design and reality, envisioned innovations and success, controlled lab environments and real-life settings.

These same gaps are also main challenges in DSR. Designs and prototypes are refined in a stepwise manner while being evaluated on a gradual path to more natural settings. Typically, the process starts with iterative evaluations in an artificial environment as long as the project is in the ex-ante stage, i.e., prior to the prototype construction (Sonnenberg and vom Brocke 2012). This means DSR faces the same challenges as digital innovation, namely the difference between testbeds and real-world settings, as well as the difference between intended use and real use in the field. The resulting two "genres" of DSR have been called the "laboratory approach" and "practice approach" (Goldkuhl and Sjöström 2018), signifying the epistemological gap between research and practice. Given the timescale required for a proof-of-use – i.e., the demonstration of *practical* value creation and value gain as the final phase of the *last research mile* (Nunamaker et al. 2015) – many DSR projects do not exceed well-designed laboratory experiments and do not reach validation in real-life settings. The absence of methodological guidance on evaluations of innovations at earlier stages has also been called the "evaluation gap" (Venable et al. 2012).

This paper presents a novel ex-ante evaluative approach for DSR to reduce these gaps called "disciplined simulation". It is derived from pretotyping ("*pre*tend pro*totyping*," i.e., a set of techniques to "test the initial appeal and actual usage of a potential new product by simulating its core experience with the smallest possible investment of time and money" (Savoia 2011, p. 21)) and puts a forecast of the proof-of-use at the beginning of the research process. This serves to determine whether users (or customers) are interested in a potential solution even before it is further developed into a proof-of-concept ("to demonstrate the functional feasibility for a potential solution" (Nunamaker et al. 2015, p. 16)) or proof-of-value ("to demonstrate that a solution can be useful for real problems" (Nunamaker et al. 2015, p. 20)). Especially when a number of competing options are on the table, our method helps to prioritize the most promising innovation opportunities and discard those less likely to engender desired usages, thereby reducing waste of time and resources in developing artifacts at later and higher maturity iterations that ultimately fail to achieve the targeted user interest and, ultimately, usage.

To provide scope and focus for this paper in accordance with our stance as defined above, we restrict ourselves to the early phases of innovation (i.e., before a prototype is developed), where the problem and solution spaces are broad, and many concurring ideas co-exist. Dealing with many approaches and ideas is a common problem for innovation frameworks. Design thinking (Brown 2008), like many other methods, encourages the initial expansion of the solution space. Afterward, a convergent phase reduces the amount of ambiguity while raising the level of prototype maturity. Serrat (2017) defines design thinking as applying an abductive logic of reasoning in a subjective, interpretative, and experimental environment. However, while we see these methods and frameworks as useful for generating creative solutions, they lack the necessary discipline to foresee if possible solutions resemble real-life situations closely enough.

Since we value creativity as central to innovation, we propose a specific method for reducing the gap between creativity and the actual use of artifacts still to be created. This may sound paradoxical if not impossible, but we claim that a disciplined approach for designing yet-to-be-prototyped services (i.e., a "disciplined simulation" of "phantoms" as per in our title) and their trial with genuine users is a better and viable way to reduce and manage the unpredictability of digital innovation and DSR and gives legs to creative ideas. Before discussing the disciplined simulation approach in more detail, we outline some philosophical underpinnings of this method. After explaining our approach, we assess its practicability to industry cases that have approached innovation this way. Finally, we conclude with a discussion of the merits and limitations of the approach.

Philosophical Underpinnings

Key philosophical underpinnings help situate our proposed approach against design science research and digital innovation more broadly. We have found that the central tenets of pragmatism are particularly appropriate for a method that reduces the gap between actions and consequences and, more poignantly here, assess objectives based on their outcomes. A pragmatist view means that nothing is abstract, and everything exists only to the extent that it has effects: "Consider what effects, which might conceivably have practical bearings, we conceive the object of our conception to have. Then, our conception of those effects is the whole of your conception of the object." (Peirce 1878, p. 293). In the context of design science and digital innovation, this view may spur polarizing reactions. Those who conceive design and innovation as ideas that succeed or not depending on their inner merit may see pragmatic approaches as not seeing the wood for the trees. On the other side, those who believe that context and practice define the fate of innovation may find it helpful to adhere to approaches that look for nothing beyond what can be realized in practice.

In other words, a pragmatist agenda offers concrete mediation between people and organizations expecting technology to make a difference to their daily lives on the one hand, and an IT industry driven by hype on the other. A pragmatist approach to information systems is about foregrounding effects and consequences as much as possible. This means testing IT in practice as early and as much as possible and constantly performing reality checks. Therefore, it could be claimed that DSR is pragmatic (literally) by design, as it mirrors the main steps of pragmatic scientific inquiry (Martela 2015, p. 545): First, an "indeterminate situation" (practice) occurs, then comes the problematization and solution-targeted altered experience (research), leading to an improved situation. Consequently, design and application can and should go hand in hand, rather than remaining conceived as distinct stages of design and development.

As Flores et al. (1988) argued compellingly already decades ago, technology is not the design of artifacts but of new practices that technology allows – which is exactly what our approach of disciplined simulation tackles. In short, by foregrounding practice, pragmatism does not downplay the creative side of innovation. Instead, disciplined simulation aims to amplify the effects of creativity by grounding it in real practice as soon as possible, so to generate more viable prospects through this process. In short, our pragmatist agenda promises to achieve that:

1. by allowing practical tests of design ideas even before a prototype has been made,
2. by producing empirical data to inform decision-making even before a budget for a prototype has been agreed upon, and
3. by generating further ideas for innovative systems that might not have been conceived before or without pretotyping.

To position our proposal in existing design practices, we continue by outlining the role of evaluation in DSR.

Evaluation in Design Science Research

Much focus has been put on the management of innovative information systems that allow for strategic differentiation and/or reduction of costs. For that, several methods and techniques have been developed to select and funnel these ideas throughout the different phases of prototyping and evaluation. Our proposal does not aim at substituting existing methods but at extending the toolkit that designers and managers can rely upon.

Turning an invention – or likewise, an "exaptation" or an "improvement" (Gregor and Hevner 2013) – into innovation (i.e., adopted by a community of practice) is riddled with obstacles. A frequently used term is the "valley of death," highlighting the many inventions that perform well during initial development but fail to succeed in practice when there are no funds for extended development (Frank et al. 1996). While there might be many (technical) barriers to sustained artifact usage or market entry ("permitting processes," "insufficient [...] performance," "liabilities," "cumbersome contracting," etc.; see Frank et al. (1996) for the full list), one of the fatal causes of remaining stuck in the valley of death is to design a product or service that nobody wants to buy or use, constituting the "wrong it" (Savoia 2019). One central observation put forward by Savoia (2011) is that people (and processes) tend to place great emphasis on whether an artifact has been built (constructed) correctly, rather than whether that artifact was the right one in the first place. Design science research makes no exceptions here; indeed, common DSR evaluation taxonomies and frameworks either implicitly (Prat et al. 2015) or explicitly (Venable et al. 2016) suggest starting with artificial evaluation settings and then moving to more natural evaluations later, even while acknowledging that evaluations in naturalistic settings are "the real 'proof of the pudding'" (Venable 2006, p. 5). Thus, at the beginning of a typical DSR project, there is little empirical support that the final artifact has any chance of success in the real world. Some scholars suggest frameworks in which the evaluation of actual system use (Sonnenberg and vom Brocke 2012) is "the last research mile" (Nunamaker et al. 2015). It is entirely understandable that no exhaustive evaluation of sustainable use or user behavior can happen before artifact completion (and therefore sustained artifact usage). However, perfection is the enemy of good here, as DSR strives to create relevant artifacts that solve relevant problems rather than polished solutions with no problem to solve or novel practice to enact. Consequently, if DSR is pragmatically oriented, proof-of-use should take precedence over proof-of-concept and proof-of-value.

In DSR, evaluations are multi-purpose instruments, aiming to prove practical impact ("relevance") and also grounding contributions to the knowledge base ("rigor") (Hevner et al. 2004, p. 84). Both research and practice must address the non-trivial question of when to evaluate an artifact and how. "Ex-ante" evaluations are performed before the artifact is instantiated, "ex-post" evaluations are carried out after its construction (Sonnenberg and vom Brocke 2012; Venable et al. 2012, 2016). The general idea of systematic evaluations throughout any DSR project starts with ex-ante evaluations and ends with summative ex-post evaluations (Venable et al. 2016). Depending on the nature of the artifact (i.e., if the problem domain is somewhat socio-technical rather than purely technical), one might diverge quickly from artificial ex-ante towards naturalistic ex-post evaluations, requiring the availability of an artifact. Venable et al. suggest such a strategy for rigorously evaluating the artifact in realistic settings "[i]f the major design risk is social or user-oriented and/or if it is relatively cheap to evaluate with real users in their real context and/or if a critical goal of the evaluation is to rigorously establish that the utility/benefit will continue in real situations and over the long run" (Venable et al. 2016, p. 82).

For ex-ante naturalistic evaluation, DSR has so far only discussed few methods, such as action research and focus groups (Venable et al. 2012). Abstract artifacts can also be subjected to ex-ante analysis, such as algorithmic complexity analysis for formal proofs and logical reasoning (Prat et al. 2015). These methods, however, are of limited use to evaluate actual user behavior that could inform an artifact's proof-of-use. To further illustrate this point, let us consider focus groups, which have gained some attention in DSR (Tremblay et al. 2010). Focus groups are non-realistic by design: the participants' collective use and discussion of early artifacts (such as designs or models) usually deviates greatly from the intended use. Also, participants are aware of the evaluation's artificial context at all times, thus, they can judge the desirability or feasibility of a solution only by predicting their own future behavior. When doing so, they tend to place too much weight on their current intentions (Poon et al. 2014), which may produce strong evaluation biases; in other words, and according to Silverman (1998), people misjudge the gap between what they mean and what they do. Therefore, care must be taken when using focus groups for such (ex-ante) evaluations; they cannot be used to find out what users will *actually* do: "There's no way that people can predict whether they will actually want to use a product, service, or feature in practice, or whether they will even be able to use it effectively." (Goodman et al. 2012, p. 142).

Generally speaking, all "proof-of-value" (PoV) and "proof-of-use" (PoU) evaluations (Nunamaker et al. 2015) require artifact construction and are ex-post in nature. This is also in line with the updated DeLone and McLean model of IS success, stating that real organizational impact requires individual benefits that emerge from satisfactory use of the system in question (DeLone and McLean 2003). The direct consequence of the current state of the art is that artifacts need to reach a certain level of (technical) maturity to be evaluated in a naturalistic setting. One implication is that the efforts of materializing artifacts at maturity levels sufficient for PoV and PoU evaluation are usually very high, which is the opposite of what we argue here. Indeed, in the early stages of innovation, ambiguity is high in the solution space, and many different variants may exist. Even more poignantly for our argument, testing the envisioned relations between users and service providers through an artifact is postponed to a late stage, when other, possibly equally viable solutions have already been discarded in the "innovation funnel" for reasons that are not empirically justified.

Therefore, we propose a disciplined method to evaluate "phantoms," i.e., simulations of systems testable by end-users before a prototype has materialized. In this way, we address the "evaluation gap" (Venable et al. 2012) by providing methodological guidance on naturalistic ex-ante evaluation in DSR and practice. It is important to note that such ex-ante evaluations are necessarily formative, since no "real" artifact exists. However, these methods can also be used in ex-post evaluations to test artifacts summatively. Combinations are also possible, as we show in case study 3 below (i.e., formative evaluations of ideas based on already implemented systems).

To quantify the materialization level of artifacts, the European Union provides a scale adapted from NASA called the *technology readiness level* (TRL) that rates artifacts iteratively from TRL 1 ("basic principles observed") up to TRL 9 ("actual system proven in an operational environment") (European Commission 2014). Thus, a TRL of 4 ("technology validated in the lab") seems to be the lowest possible TRL for PoV evaluations. However, this scale is intended to rate complete systems (or system components), and it has been questioned whether TRLs could be safely used to rate general innovation project maturity (Héder 2017). Therefore, our disciplined simulation approach needs to provide a selective maturity model to manage (i.e., minimize) design and implementation efforts before prototyping, while keeping artifact maturity at a sufficient level to permit evaluative rigor.

Finally, some domain delimitation is required. In their taxonomy of typical DSR evaluation patterns, Prat et al. (2015) categorize "simulation-, metric-based benchmarking" and "simulation-, metric-based absolute evaluation" of artifacts as experimental evaluation techniques. While the taxonomy does not explicitly differentiate methods in terms of research phase or time of application, the simulation-based techniques are characterized as being applied to evaluate instantiations of the artifact, which are typically ex-post by nature. Also, and crucially here, such techniques are performed without actual end-users.

Lukyanenko and Parsons (2020) propose Weick's (1989) concept of "disciplined imagination" as a way to lessen "design theory indeterminacy" in DSR, i.e., the gulf of theoretical propositions and concrete practical issues. Weick (1989) describes disciplined imagination as part of theorizing; he poses that theory building processes offer several "choice points," where "theorists can act differently and produce theories of better quality" (Weick 1989, p. 519). These are the problem statements (assumptions, representations), the formulation of "thought trials" (number and heterogeneity of trials), and the selection criteria among these trials (consistency and diversity of criteria) (Weick 1989, p. 529). Our approach of disciplined simulation has significant similarities with these principles. On one side, we want to provide practitioners with early "choice points" to produce artifacts with better chances of actual use or, at least, to "fail fast" and prioritize other options. On the other side, researchers can produce real data on still imaginary solutions. Thereby, disciplined simulation facilitates evolutionary variation, selection, and retention of promising ideas, which are validated by concrete experimental evaluations besides pure "thought trials."

Disciplined Simulation as an Ex-ante Evaluation Approach

This section describes the different components of the disciplined simulation approach, which we later discuss in relation to digital innovation and DSR. Here, we argue how DSR can integrate disciplined simulation as a means of evaluation of proof-of-use, how innovation maturity levels can be used to track progress, and how disciplined simulation can be used to extend common DSR process models.

Pretotyped Proof-of-Use Evaluations

Objective: To apply pretotyping techniques to anticipate proof-of-use as early as possible and repeat that approach throughout the project.

The discussion on maturity or "fidelity" of prototypes is several decades old. Many of the basics were established as early as 1989 (Virzi 1989). Deliberate reduction of the number of features or their functionality to enable cheap prototyping was conceptualized at that time (Nielsen 1989). Back then, the idea was to reduce both features and functionality to a bare minimum sufficient to run some "simplified thinking aloud" tests (Nielsen 1989). Similar ideas can be found for service scalability (Hanseth and Lyytinen 2016). However, it also works the other way around, starting with scenarios and then expanding the solution space and maturity of the components. Scenario-based development is a frequently used technique to accomplish this (see Rosson and Carroll 2002).

More recent approaches aim to reduce the maturity level even more, to the point where there is not even a prototype. This so-called pretotyping suggests a range of techniques on various levels of maturities (see Table 1). The higher-level techniques (e.g., cardboard mockups or Wizard of Oz prototypes) also have a long tradition in (software) interaction design (Rogers et al. 2011, p. 390).

The pretotyping methods target different levels of maturity from basically non-existent (fake door) to a sophisticated level (minimum viable product, i.e., a version of the product that is complete enough to demonstrate the value it brings to the user, while requiring a minimum of effort (Moogk 2012)). However, the interesting idea behind these forms of evaluation is not only a reduction of artifact instantiation effort but rather a forecast into later evaluations from the perspective of DSR. Fake doors are an extreme example, because before anything is designed or built, one can forecast actual user interest in a product or service idea: will users "in the wild" be interested enough to click on a web ad, i.e., knock on the front door (Savoia 2019, p. 104)?

In contrast, traditional DSR aims to iteratively validate the stages of artifact development by performing (1) an evaluation of the problem, (2) an evaluation of a design, (3) an evaluation of the construct/artifact, and (4) an evaluation of the actual use (Sonnenberg and vom Brocke 2012). Pretotyping offers a forecast for the evaluation of use, bypassing the other steps almost entirely. Nevertheless, we do not see pretotyping methods as a replacement for DSR methods, but rather as an extension of its toolkit for additional early evaluations (Figure 1). While the inexpensive nature of these methods favors their use in ex-ante evaluations, their application is by no means limited to them. As we show in case study 3, these methods can also be applied in ex-post settings.

Fake Door	Provide a fake "entry" (e.g., button, advertisement) to a product/service to test interest (e.g., how many people will click the advertisement).
Façade	Provide a fake product (e.g., on a website) and a call to action (e.g., purchase) to test demand (e.g., how many people will explore the website / make a purchase).
Pinocchio	Functionless physical artifact (e.g., paper/cardboard prototype) to validate the form and fit of a product.
Re-Label	Use existing other technology while adding just some extensions and promote it under your own label.
Mechanical Turk / Wizard of Oz	"Functionality" of an artifact is provided by (concealed) human intervention.
MVP	Minimum viable product / service that can be used and sold.

Table 1. Selection of Pretotyping Methods (Savoia 2011, 2019) Deemed Applicable to IS

In terms of maturity, it is difficult to classify these early, "intermittent" artifacts within the TRL framework, as they are unlikely to show any readiness at all (except for an MVP or Re-Label pretotype). On the one hand, an MVP matches TRL9, as it is successfully sold to customers; on the other hand, it could be argued that the TRL decreases as soon as additional (currently missing) features and functions are added or major changes are made. As we can see, a change in requirements may also result in a change in the maturity level. Since such a classification is not stable, we need another way to describe the maturity of the artifact, as elaborated in the following section.

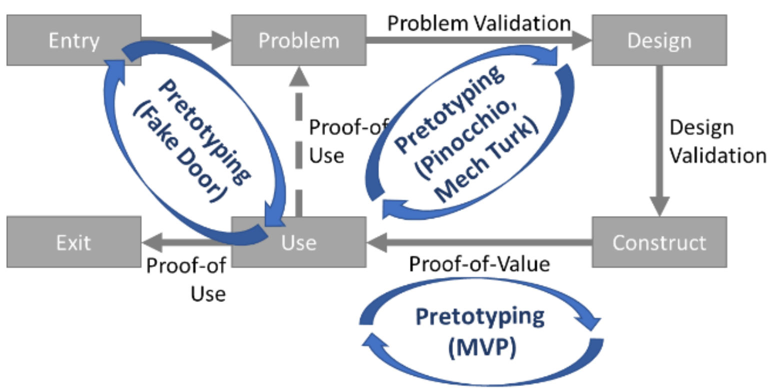

Figure 1. Adapted and Extended Evaluation Model by Sonnenberg and vom Brocke (2012)

From "building it right" to "building the right it"

Objective: To describe the innovation maturity level as a consequence of pretotyping steps performed, independent of technical maturity.

Taking technical maturity out of the equation brings us away from levels of "building it right" towards levels of building "the right it" (Savoia 2019), while still maintaining the ability to track progress. Where possible, the evaluations relevant to an increase in maturity should happen in a naturalistic setting to increase the external validity of results. This contrasts with the classical notion of progressive evaluations in DSR (e.g., see Figure 2 in Venable et al. (2016, p. 80)), which begin with evaluations in artificial settings and progress to more naturalistic ones. However, it must be noted that evaluations using pretotyping methods approaches do not necessarily test the artifact itself (most notably fake doors and façade), but rather the appeal and viability of the envisioned relation between users and service providers, through a yet-to-develop artifact (Table 2).

Level	TRL	IML	Time	Type
0	-	Idea generated	Ex-ante	Naturalistic
1	Basic principles observed	Fake door entry / façade created and offered		
2	Technology concept formulated	Fake door entry / façade validated		
3	Experimental proof of concept	Pinocchio created		Artificial
4	Technology validated in lab	Pinocchio validated		
5	Technology validated in relevant environment	Mechanical Turk, re-labeled product constructed	Ex-ante/ Ex-post	Naturalistic
6	Technology demonstrated in relevant environment	Mechanical Turk, re-labeled product sold		
7	System prototype demonstration in operational environment	MVP constructed and offered	Ex-post	Naturalistic
8	System complete and qualified	MVP sold and supported		
9	Actual system proven in operational environment	MVP transferred into continuous integration / continuous delivery scheme		

Table 2. Comparing TRL to Innovation Maturity Level (IML) and Adding Classification (Ex-Ante/Export and Naturalistic/Artificial from Pries-Heije et al. (2008))

Integrative Process Cycles

Objective: To design pretotyping projects through a disciplined process that incorporates development and evaluation steps. The process begins with a pretotyping phase, followed by a prototyping phase, and ends at IML9 with an MVP.

First, back to DSR: Hevner et al. (2004) describe DSR as a paradigm featuring rigor and relevance activities to ground the design in the field as well as in the scientific knowledge base. Specifically, the rigor cycle foresees “additions to KB [knowledge base]” (Hevner 2007, p. 88). One could argue that an organization innovating information systems is not necessarily interested in adding knowledge to the scientific community. Peffers’ et al. popular DSR methodology (Peffers et al. 2007) describes practical DSR projects with the following phases: (1) problem definition, (2) solution objectives, (3) design, (4) demonstration, (5) evaluation, and (6) communication. Following the previous line of thought regarding organizations in practice, phase 6 could be skipped in terms of scientific publications, but certainly not concerning internal communication and preservation of knowledge. Furthermore, innovation (rather than design science) does not necessarily have to address “unsolved and important business problems” (Hevner et al. 2004, p. 84). While problem-solving and entrepreneurial innovation are not mutually exclusive (i.e., some innovations *do* solve existing problems), problem-solving is not always necessary, as it is possible to create and fulfill a demand without there being a salient problem beforehand (e.g., the introduction of smartphones to the general population). Hence, our disciplined simulation approach is less strict on problem definition (though not to be confused with relevance!) and formal requirements for communication. Figure 2 shows the different activities and their dependencies. In the following section, the two main phases are examined more closely.

Phase 1 – Iterative Pretotyping (IML 0-6): This phase is characterized by the idea of keeping an innovation constantly in an evaluation loop. In contrast to DSR practices, where the first task is to define the problem and prove relevance, disciplined simulation requires initially evaluating the level of demand for a solution in a naturalistic setting. In this way, the innovation idea remains in the evaluation loop from the very beginning (starting as a "phantom") – only the materialization of the artifact (that instantiates the idea) changes over time.

From the outset, the artifact is purely virtual and therefore not tangible in any way. As the project progresses, more and more parts of the artifact reach increasingly higher levels of maturity. At Levels 5 and 6, only the interface of the artifact (towards its users/customers) needs to be instantiated (see the discussion of philosophical underpinnings), while the rest can be simulated. At IML 6, the demand for a tangible/usable product or service is established, and the first set of functional requirements from the customer are defined and evaluated.

During these initial steps, disciplined simulation can and should be grounded in the scientific knowledge base. Depending on the level of maturity, different ways of grounding can be applied. Goldkuhl (2004) describes "conceptual grounding," "value grounding," and "explanatory grounding" as additional grounding strategies that can (and should) be applied in parallel to "empirical grounding." Although his focus is on grounding prescriptive statements of design knowledge, we argue with the following case studies that knowledge about the viability of innovation ideas can be captured/sensed empirically earlier than previously assumed. Following along, even if a pretotype fails to show a positive correlation between actions and goals, it may disprove (or falsify) viability in a naturalistic setting.

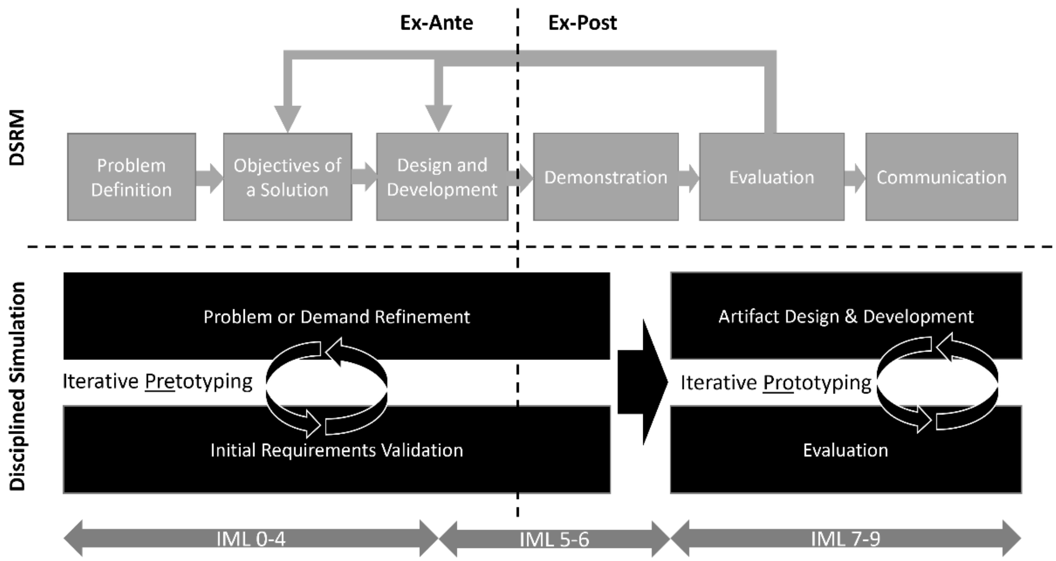

Figure 2. Disciplined Simulation Phases Adapted from DSRM (Peffers et al. 2007)

Phase 2 – Iterative Prototyping (IML 7-9): This second phase is much more traditional and follows the decades-old idea of a spiral model (Boehm 1988) to manage risk and uncertainty during software product development. Today, comparable iterative approaches would probably fall into the domain of agile software development (see Fowler and Highsmith 2001) and can be supported by continuous integration (CI) and continuous deployment (CD) pipelines (see Shahin et al. (2017) for an overview). Further, all standard evaluation methods can be applied at this stage. Finally, if IML9 is reached, the product has been successfully introduced to the market in MVP status. Applying CI/CD enables a seamless transition from the research innovation process into traditional software product development. Keeping the actual usage in the field in focus should lead to a higher success rate of fully developed prototypes. The "right it" has now been instantiated.

Application in Practice

This section illustrates how disciplined simulation can be applied by describing three industry cases conducted by one of the authors. This multiple-case study approach enables "[…] researcher[s] to explore differences within and between cases" (Baxter and Jack (2008, p. 548) following Yin (2003)). It also allows different levels of "grounding" (Goldkuhl 2004): besides "conceptual" and "value grounding," we can generate early "empirical grounding" that supports our approach.

The cases were conducted applying pretotyping methods before their theoretical conceptualization in relation to DSR. So, they informed the components described in the previous section. As practical examples of applying pretotype methods, they provide material to illustrate the benefits and challenges of our approach retrospectively. Case studies 1 and 2 showcase innovation testing in Phase 1 (iterative pretotyping), using artifacts at IML1, i.e., using combinations of fake doors and façades. At the beginning of case study 3, the innovation project had already been developed to IML7 (MVP constructed and offered). As the MVP had not been successful, however, pretotyping was restarted at IML1 – but using the technological base of the MVP.

Although participants did not actively consent to participating in the evaluations (given the nature of the pretotyping methods used), it is unlikely that they could have been harmed by these evaluations. On the one hand, the companies involved supported the corresponding efforts and intended to use the results to initiate fully-fledged services at a later stage; on the other hand, the privacy of users was protected. Also, participants who wanted to purchase the respective services were informed about the experiment (i.e., the service did not exist but was under development) and were asked to fill out a questionnaire, including open-ended feedback.

Case 1: Payment App

Background: The following case is based on an innovation project in a Swiss bank performed within a dedicated innovation lab. The starting point of the project was the observation of one of the lab's innovators that – given all the available consumer-oriented technology – paying household bills was still not very straightforward. Various forms of invoice (on paper, with or without payment slips, digital via email, or e-bill via electronic banking) met different processes of payment (physical bank counter, physical ATM-like payment machines, manual transfer of payment data into electronic banking software, automated scan features in mobile apps, direct debiting, etc.), which were not compatible with each other or integrated into a standardized payment process.

Unsurprisingly, from various conversations the innovator learned that most people seemed to find collecting and paying bills typically once a month a tedious process, prone to errors due to forgotten or lost documents. Furthermore, the innovator discovered that these people would welcome having the chore taken away from them and, crucially, would be willing to pay for such a service. This would have allowed the bank to strategically differentiate itself in payments, which had been heavily commoditized in recent years. At the same time, it was realized that many services designed to increase customer convenience in banking were underused, especially in electronic banking.

Approach: From this starting point, the innovator decided to conduct a study to deepen the understanding of (1) what customers expect regarding convenience, (2) what current services they do (or do not) use and why, (3) how to provide greater convenience with making payments, and (4) how much customers would be willing to pay for such a service.

The innovator knew that the first two aspects could be readily and cost-effectively addressed by standard desk research, analysis of the bank's existing customer (behavioral) data, and customer surveys on their usage and appreciation of payment options. Concerning the second two questions, she knew that a solution could be prototyped and tested with customers. However, she also knew that this would involve a lot of effort and expense and take longer than senior management would authorize. Worse still, the approach would cause problems in terms of validity, as the evaluation of a prototype in an artificial setting (interviewing customers, organizing focus groups, or even controlled experiments) would reveal little about whether the solution would withstand the rigors of the outside world.

Together with an external consultant (one of the authors), she decided to take a different approach: pretotyping. The goal was to answer the questions using minimal time and resources – performing desk research and analytics, administering an online survey as well as providing a testable solution concept within two weeks, and building the pretotype artifact in parallel (also for a maximum of two weeks), before testing the solution in a "live" scenario for a total of four weeks.

Solution: From the desk research, analytics, and online survey (61 participants from personal networks), most initial assumptions were confirmed, and specific requirements for the potential solution were derived (e.g., key features and a price range). The solution concept was prepared in a few days, involving subject-matter experts and business managers. It consisted of a smartphone (or desktop) app that sent all invoices to a central banking site (via photo, email, or physical post). Then at the end of each month, the app would present the user with an overview of pending payments and allow approval with one click.

The solution was implemented through an advertising agency using the pretotyping methods "fake door" and "façade", by placing online advertisements (via Google and Facebook) that led to a website describing the (not yet existing) service and explaining the three different service levels on offer. As a proxy for the actual purchase decision, people trying to purchase the service were asked for their name and email address in the first step of the checkout process. Importantly and to keep the purchase situation as realistic as possible, potential customers were not made aware beforehand of the website's true nature and were not informed about this until they had attempted to purchase the service and had provided their personal information. Only at this point were they asked to participate in a survey.

Results: Before testing the pretotype in the field, hypotheses of the main outcomes were defined, specifically regarding the conversion rates (CVR) of visitors viewing general information on the site (conversion 1; hypothesis: 10% of visitors), progressing to view detailed information (conversion 2; hypothesis: 5% of visitors) and/or trying to purchase the service (conversion 3; hypothesis: 1% of visitors). All hypotheses were defined based on the agency's experience from similar projects in the same industry.

Figure 3 shows the results after four weeks with roughly 349,000 impressions (number of times the advertisements were displayed) and 2,900 website visitors. All expectations for conversion rates were exceeded, and it was concluded that the innovation idea worked in principle and justified further development.

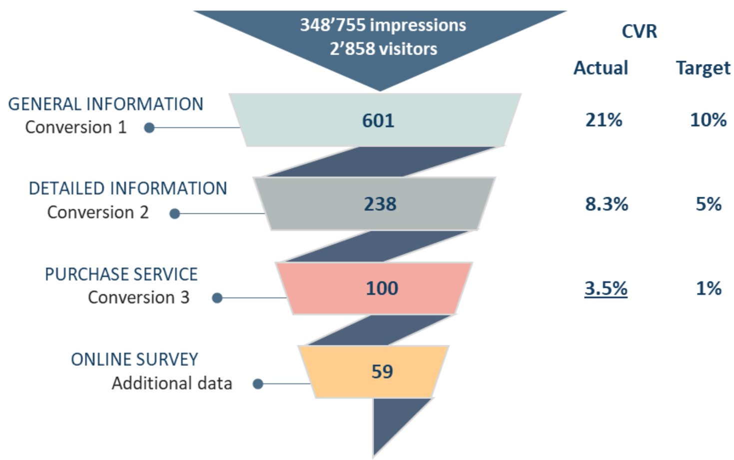

Figure 3. Pretotyping Results of Case 1

Further development: As was customary at the bank for innovation projects (following a traditional innovation funnel), the project results were presented to a panel of innovation delegates of the different functional departments, and the idea was deemed worth pursuing. According to the bank's innovation process, the idea would have had to be further developed and financed by the department concerned (i.e., client solutions/payments). However, at the time of the project, the department was already busy with other initiatives for the coming few years and did not want to adopt the project. Overall, the project was still felt to have been a success because it achieved its goal of realistically testing the innovation idea *before* implementing possible solutions, which proved valuable given the organization's subsequent lack of interest in further development.

Case 2: Website for Self-Service Pension Provision

Background: This case illustrates an innovation project by a Swiss bank. One of the most critical concerns among the Swiss population is pension provision and risk coverage. For this reason, the bank decided to include an innovative service into its midterm business strategy.

The project team had already created rudimentary personas of the target customer segments as well as basic customer journey maps (defining the expected goals, behaviors, and desired service experiences). It had also developed different foundational service components of an envisioned self-service portal with advisory support.

Although the strategic budget was available and the internal IT was interested in the implementation, the project manager was hesitant to start building a complex and expensive online platform directly: Were the previous assumptions correct and would there be a genuine demand for the envisioned solution?

Approach: While some preliminary work had already been done, the following questions were still unanswered: (1) What were the specific customer needs, (2) would the service convince potential customers, and (3) what would be an attractive starting point (minimum viable product)?

With the external support of one of the authors, the project manager decided to tackle these questions using pretotyping methods. As a starting point, the first question was addressed by desk research on existing studies and services concerning pension provision and risk coverage. Then, to learn more about customer problems, needs, goals, and the challenges of existing and potential future services, the project team conducted six pilot interviews and an online survey via social media and personal networks with a total of 81 participants.

These research findings were used to refine the customer segments, supplement the previously developed service ideas, and translate them into a pretotype service to test customer demand and willingness to pay (Question 2 above). Based on the pretotype findings, the service components would then be detailed and prioritized into an MVP, i.e., a minimum viable online portal (Question 3). To limit time and resources, fact and need-finding (desk research, interviews, and the survey) were limited to four weeks, with the last week overlapping with a two-week concept and design phase for the pretotype, followed by a two-week implementation and 4-6 weeks testing in the field. A team of experts (subject specialists on pension provision and risk coverage, product management, marketing) was involved in all activities.

Solution: Based on the findings from desk research, interviews, and an online survey, four general customer types were identified – two of which were considered interested in the envisioned online, self-service product. Based on these two types, the project team developed three personas with differing sociodemographic backgrounds, subject and online affinities, behaviors, and expectations regarding self-service and advisory.

For the pretotype evaluation, Google and Facebook advertisements (fake doors) and a website for the product ("façade") were implemented in cooperation with an external advertising agency. The advertisements were also displayed on the logout page of the bank's online banking system.

The pretotype website contained high-level descriptions of the service – what it is and how it works – and detailed topical information regarding pension planning, risk coverage, wealth building, and estate planning. It also featured a call-to-action to register with the service, collecting personal data such as name, phone number, and age in the first stage of the checkout process. To further verify willingness to pay (roughly obtained from the interviews and online survey), the service registration was (randomly) presented from three pricing variants, including a free option for the self-service product without advisory services. To create a realistic decision-making situation, visitors to the website were not made aware of its experimental nature until they had attempted to register with the service. Visitors who registered with the service were then informed that the service was not yet available and asked to participate in a short survey.

Results: The pretotype was tested against several propositions regarding the visitor conversion rates (CVR). Based on the participating agency's experience with similar projects in this industry and given the importance of the topics and their challenges for individuals, it was assumed that more than 30 percent of the website visitors would in some way interact with the information presented (e.g., using the navigation menu; Conversion 1). As no reliable data were available, no assumptions were made regarding the distribution of clicks on detailed topical information (Conversion 2); however, more than 2.5 percent of website visitors were expected to register with the service (Conversion 3).

Although overall interest in the service offering was considerable (with over 10,000 persons visiting the website based on the advertisements), the expected activity could not be achieved (Conversion 1: 25.5% < 30%; see Figure 4), and only a few visitors (1.8%) sought out the detailed topical information (Conversion 2). In total, only 0.4 percent attempted to register with the service, which was much less than the expected 2.5 percent (Conversion 3).

Detailed information on the individual topics was displayed with approximately equal distribution, except for estate planning, which was accessed infrequently. Due to the technical implementation, relationships between selected topic and service registration could not be established.

The results on willingness to pay were inconclusive; owing to the limited number of active online visitors, the variants were not displayed in an equally distributed manner and did not show significant differences. However, at face value, the registering visitors slightly preferred the free, self-service-only variant; the most expensive variant was tried to purchase least frequently.

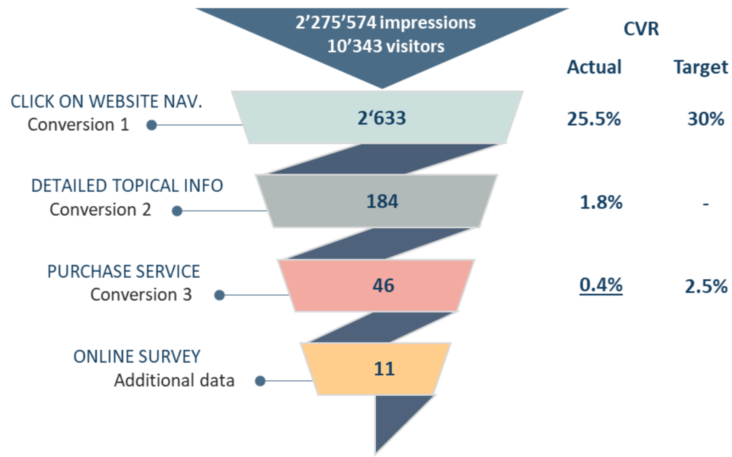

Figure 4. Pretotyping Results of Case 2

Further Development: Although all the assumptions regarding conversion rates were disproved by the results, it was decided to continue the project. There were two main reasons for this. First, it was assumed that the number of visitors already indicated an underlying interest in the topic. Second, and more importantly, the initiative had to be followed up because of the bank’s strategic focus. Therefore, it was decided to develop the concept into detailed customer journeys and designs of the target solution further as a next step. After promising results in this second phase, with detailed requirements for an MVP, the leading management supporter of the project left the company, at which point further activity was suspended.

Case 3: Personalized Skin Care

Background: In this case, we present the course of a strategic innovation project in the cosmetics industry. The company described is a manufacturer in the cosmetics private label business, i.e., in manufacturing products for other companies’ brands.

To respond to current trends in the industry (shift from offline to online, personalization of products, services and packaging, new types and forms of cosmetics application, etc.) and differentiate itself with own brands in the direct-to-consumer market, the company developed the vision of a new business model: a personalized premium cosmetics line marketed and sold exclusively online. The basic concept was as follows: on a website, the customer was asked to fill out a detailed questionnaire about her skin needs and upload corresponding photos of relevant skin areas. Cosmetic experts analyzed these inputs and created an individual formulation of a suitable skin cream. Based on the analysis, the customer could order her individualized product directly (and repeatedly) via the website’s online shop; in case of changes (skin needs, seasons, etc.) the customer could further optimize her products’ formulas.

As a traditional manufacturer of mass-produced products for other brands, the switch to personalization and online sales were particularly challenging for the company. After concept development with substantial market and customer research, brand and package design, as well as development of basic cream formulas, the company decided to test its offering using an MVP approach. While the website and its online shop were fully functional, the skin analysis, compilation of individual formulas as well as production, packaging and shipping were performed manually. Despite the complex development process, the MVP did not achieve the expected sales. Therefore, the project team decided to investigate the possible causes in more detail.

Approach: When he was brought into the project, the participating author recommended to re-validate the product's main value proposition using pretotyping methods. With the support of an advertising agency, the project team decided to proceed in two stages: (1) learning about the product's problems (i.e., traffic vs. conversion rates), (2) testing whether conversion rates could be optimized. A maximum duration of two months was set for each stage.

Solution: In order to accurately track customer interaction on the website, various measurement points were placed on the existing infrastructure. For both stages, dedicated advertising campaigns were launched to direct customers to the website and monitor their behavior. While the first stage was mainly based on the original website and contained only a few "quick fixes" regarding interaction, the second stage involved substantial changes to the site.

It should be noted that, technically speaking, the existing product infrastructure was not a pretotype – after all, the advertised product was already being manufactured. In terms of their objectives, however, the experiments were nonetheless pretotyping experiments, as they focused on the value proposition of the product (customer interest as conversion rates) rather than the product itself (formulas, design, etc.). To that end, the experiments could also have been conducted without the actual product.

Results: The pretotypes were tested against long-term target conversion rates based on the agency's experience with similar projects in the same industry, with an estimated hurdle for profitability at a purchase rate of 1%. Despite only minor design adjustments in the first stage, all conversion rates improved significantly (Figure 5, Stage 1). Nevertheless, overall conversion rates remained insufficient, with a purchase rate of only 0.06%. Results showed that the conversion rates were negatively impacted by the design of the sales funnel.

An analysis of user behavior revealed several reasons: the skin analysis questionnaire was too long and greatly weakened the conversion rate. This was also true for the mandatory photo upload. More importantly, it was found that the price of the product kept many customers from buying – since the product was in the premium range and correspondingly expensive to manufacture, this had a strong impact on the potential profitability of the offering. Based on these findings, various adjustments were made in the second stage. To clearly assess their effectiveness in context, the adjustments were tested iteratively. Changes included a significant shortening of the questionnaire (verified with A/B tests), making the photo upload optional, as well as significantly reducing the price point using discount codes. Again, the adjustments yielded significant improvements in conversion rates (Figure 5, Stage 2). Overall, however, the purchase rate fell far short of expectations – and at a greatly reduced price point, which severely undermined profitability of the product.

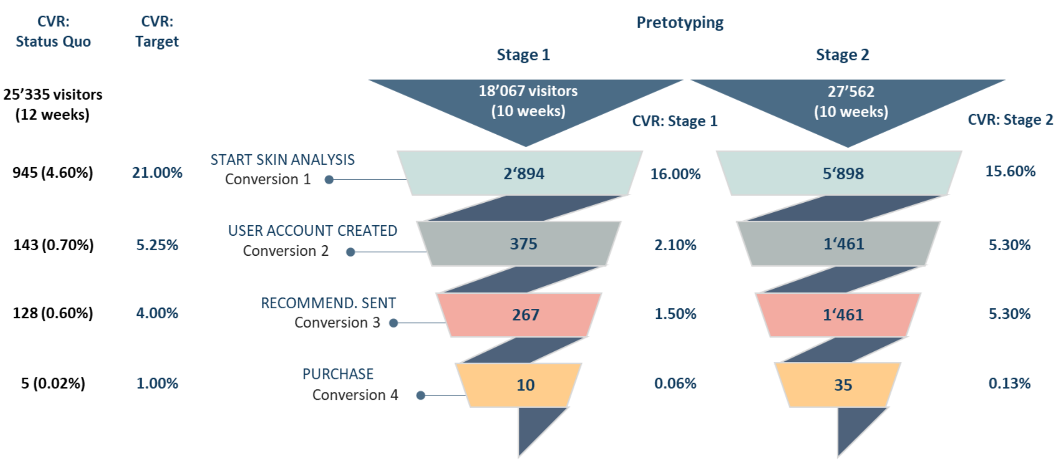

Figure 5. Pretotyping Results of Case 3

Further Development: After the second phase, the project team was motivated to test further optimizations, especially how to attract repeated purchases, which were economically more interesting (all previous purchases were made by first-time buyers, none of whom bought again). Given the results, however, the management decided not to continue with product development.

Discussion and Final Remarks

In our research on digital innovation, we have sought to bridge the gaps between the laboratory and practice approach of DSR by introducing a practical ex-ante evaluation approach derived from pretotyping. The case studies presented exemplify how the three components of disciplined simulation can be applied in practice:

(1) Pretotyped proof-of-use evaluations: The cases show how pretotyping can be used from the outset of the innovation process to test the presumed future application of an idea. With minimal, timeboxed problem verification (desk research, interviews, online surveys), the pretotypes were designed, constructed, and tested at the earliest possible moment, thereby accelerating or even bypassing steps of the problem-design-construct-use loop. Indeed, the projects of the cases 1 and 2 designed, built, and evaluated pretotypes (façades with fake doors) in a matter of a few person-days, allowing the project team to estimate potential customer interest and, by proxy, presumed future use. Case 3 shows that such proof-of-use evaluations can also be conducted iteratively and, more importantly, can also be applied to revisit assumptions of value propositions even in later stages of an innovation project (e.g., when an MVP does not meet its goals). As it is expanded upon later, it is important to anticipate that – even if the hypotheses tested here were not derived from theories – disciplined simulation holds potential to be deployed also for theory development, to the extent that pretotypes are derived from specific conceptualizations to be evaluated in practice.

In general, disciplined simulation of innovative ideas for their early evaluation through "phantoms in the real world" also promotes earlier decision-maker involvement and decisions based on real-world data so obtained. In other words, our phantoms produce real data of imaginary innovations. Indeed, the practical nature of pretotyping methods lends itself well to being conducted in close contact with decision-makers, as this phase of innovation often focuses on strategic issues (e.g., is there a market for the idea, is it economically viable, and does it fit in the company's service or product portfolio?). While working with senior management on operational issues can be challenging (time restrictions, balancing different opinions, sense-making of results, etc.), such management involvement can provide the necessary legitimation and thus support for a project. For example, in cases 1 and 2, the seniority of the person in charge enabled the project to be funded and launched quickly. As we can see in case 2, however, involving senior managers can also lead to a quick shutdown of the project – for example, when the driving force leaves the company. Using pretotyping methods in close coordination with management can avoid letting project ideas fall into the "valley of death" because they are addressing "the wrong it." This could arguably also have been achieved for the project in case 3 if its value proposition had been evaluated earlier using pretotyping, which could have avoided large investments in product infrastructure development.

As can be seen from the cases, disciplined simulation does not necessarily lead to more successful innovation – in fact, in all cases the projects were discontinued. On the contrary, the goal of disciplined simulation is to discard innovation projects as early as possible, thus preventing the waste of time and resources. Even if projects are stopped for reasons other than lack of proof-of-use (cases 1 and 2), the financial damage is arguably less than if a traditional prototyping process had been followed. Case 2 shows that even a failed proof-of-use does not necessarily lead to a project being discontinued, as there were other reasons for continuing. However, here the lack of proof was explicitly accepted in a pretotype-informed decision.

(2) From "building it right" to "building the right it": In our case studies, the selection of pretotyping methods was not guided by technical considerations but by the (lack of) maturity of the idea, which was very low at the beginning. The use of a pretotyping approach allowed the project teams to select the evaluation method with the lowest requirements for artifact details that still allowed them to gauge the interest of potential users. While prototyping would have required a lot of readily available information (e.g., requirements and customer needs) to build and test the artifact, pretotyping did not force the project teams to decide on details and functionality of artifacts they still knew very little about. It is important to recognize that reconsideration of maturity is also possible if an idea has developed in a less than promising direction (case 3).

A focus on the maturity of the idea rather than technology also helped in the interaction with the IT department in case 2. The gathered user data had to be stored within the bank's IT systems to be legally compliant, which required a software interface to the pretotype. With the pretotype perspective and its focus on speed and efficiency, it was possible to focus on a temporary (throwaway) solution and avoid the lengthy integration coordination with the overall IT architecture had the team built an actual prototype or pilot of the target solution.

(3) Integrative process cycles: An important feature of the case studies was that the stakeholders knew that it was an inherently iterative approach to test different versions of the idea and make iterative adjustments within the evaluation of a single version. Only when the teams felt they could not gain any further insights at the same maturity level was the next IML targeted. In case 2, the team discussed whether to carry out another pretotyping iteration to gather additional data, as another version of the pretotype could have been evaluated at a fraction of the cost of the first iteration (since the setup and infrastructure were already available). However, as the team did not expect substantial gains in knowledge from another iteration, it decided to move to the next maturity level, i.e., extend and detail the potential solution with requirements engineering towards a minimal viable product (MVP). As shown in case 3, such cycles can also be performed "backwards" with respect to the IML, if a developed solution does not achieve the desired goals.

Our approach is best suited for formative, ex-ante evaluations, i.e., before an IT artifact is created. This will usually be required in the exploratory stages of the research or innovation process, when different directions of development of an artifact are emerging. However, as anticipated in case 3, the approach can also be applied to existing IT artifacts to revisit earlier Innovation Maturity Levels. Thus, the use of pretotyping is conceivable for theory-driven design science artifacts, to the extent that competing pretotypes are designed based on different theoretical assumptions or options. Of course, this would not substitute established ex-post (summative) evaluative methods with respect to validity and reliability. Disciplined simulation simply expands the tools available to researchers and practitioners (especially designers and managers). Thus, we suggest that researchers can benefit from disciplined simulation for the same reasons as practitioners, i.e., increasing relevance of the process outcomes by eliminating early ideas that would not work in practice.

In sum, what are the consequences of pragmatism in digital innovation research and DSR? The main implications of our work are:

1) well-grounded empirical studies of practices can start earlier than is currently the case,
2) exploratory studies can deploy disciplined simulation beside focus groups and before prototypes are available,
3) close engagement between researchers and practitioners early on can be based on more solid data than usual brainstorming techniques,
4) constant focus on use and practical consequences of innovation ideas should affect designs, plans, and development iterations,
5) context-dependent insights are less generalizable, but more actionable.

Finally, the potential limitations of our research should be mentioned. While based on actual industry practice, our case studies provide preliminary evidence only for parts of our approach, i.e., Level 1 of our innovation maturity level (IML) model and Phase 1 of the suggested integrative process cycles. Therefore, we cannot offer conclusive empirical evidence that proceeding to higher IMLs and later stages of the innovation process (especially transitioning to traditional approaches) will deliver all the discussed benefits of the approach. Furthermore, due to the natural settings it targets, our disciplined simulation approach raises questions regarding validity, which can be considered from two perspectives: the validity of the method and the validity of the artifacts (pretotypes). Since disciplined simulation is essentially an experimental method, internal validity is especially important, i.e., the "question of whether the observed results were actually caused by the experimental treatment instead of by something else" (Briggs and Schwabe 2011, p. 98). Evaluating "simulated" artifacts in naturalistic settings ("in the wild") poses great challenges to such validity, as many co-causes are difficult to discern: participant behavior depends on several unobservable variables that cannot be measured or controlled; in addition, the spending and display algorithms of the ad platforms used (Google and Facebook) may be constantly and dynamically changing. This also negatively influences the experiments' generalizability: the specific results are so strongly linked to the time of execution that conducting the same experiment at different times is likely to lead to different results. We conclude by stressing that for ex-ante evaluation, the practical validity of "simulated" artifacts in naturalistic settings is nevertheless higher than in laboratory settings; they allow gathering real-world usage data rather than abstract user preferences identified using, e.g., focus groups. Thus, this can also challenge the "groupthink" that designers and managers may develop because of some initial enthusiasm.

References

Baxter, P., and Jack, S. 2008. "Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers," *The Qualitative Report* (13:4), pp. 544–559.

Boehm, B. W. 1988. "A Spiral Model of Software Development and Enhancement," *Computer* (21:5), pp. 61–72.

Briggs, R., and Schwabe, G. 2011. "On Expanding the Scope of Design Science in IS Research," in *Proc. DESRIST 2011* (Vol. 6629), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 92–106. (http://www.springerlink.com/content/21q1106611nw173k/).

Brown, T. 2008. "Design Thinking," *Harvard Business Review* (2008/06). (https://hbr.org/2008/06/design-thinking).

DeLone, W. H., and McLean, E. R. 2003. "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems* (19:4), pp. 9–30.

European Commission. 2014. *Technology Readiness Levels (TRL) - Extract from Part 19 - Commission Decision C(2014)4995.* (https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf).

Flores, F., Graves, M., Hartfield, B., and Winograd, T. 1988. "Computer Systems and the Design of Organizational Interaction," *ACM Transactions on Information Systems* (6:2), pp. 153–172. (https://doi.org/10.1145/45941.45943).

Fowler, M., and Highsmith, J. 2001. "The Agile Manifesto," *Software Development* (9:8), pp. 28–35.

Frank, C., Sink, C., Mynatt, L., Rogers, R., and Rappazzo, A. 1996. "Surviving the 'Valley of Death': A Comparative Analysis," *The Journal of Technology Transfer* (21:1–2), pp. 61–69. (https://doi.org/10.1007/BF02220308).

Goldkuhl, G. 2004. "Design Theories in Information Systems - A Need for Multi-Grounding," *Journal of Information Technology Theory and Application* (6:2).

Goldkuhl, G., and Sjöström, J. 2018. "Design Science in the Field: Practice Design Research," in *DESRIST 2018 - Designing for a Digital and Globalized World*, pp. 67–81. (https://doi.org/10.1007/978-3-319-91800-6_5).

Goodman, E., Kuniavsky, M., and Moed, A. 2012. Observing the User Experience: A Practitioner's Guide to User Research, Elsevier.

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337–355. (https://doi.org/10.25300/MISQ/2013/37.2.01).

Hanseth, O., and Lyytinen, K. 2016. "Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet," in *Enacting Research Methods in Information Systems*, Springer, pp. 104–142.

Héder, M. 2017. "From NASA to EU: The Evolution of the TRL Scale in Public Sector Innovation," *The Innovation Journal* (22:2).

Hevner, A. 2007. "A Three Cycle View of Design Science Research," *Scandinavian Journal of Information Systems* (19:2). (http://aisel.aisnet.org/sjis/vol19/iss2/4).

Hevner, A. R., March, S., Park, J., and Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly* (28:1), pp. 75–105.

Lukyanenko, R., and Parsons, J. 2020. "Design Theory Indeterminacy: What Is It, How Can It Be Reduced, and Why Did the Polar Bear Drown?," *Journal of the Association for Information Systems* (21:5), pp. 1343–1369. (https://doi.org/10.17705/1jais.00639).

Martela, F. 2015. "Fallible Inquiry with Ethical Ends-in-View: A Pragmatist Philosophy of Science for Organizational Research," *Organization Studies* (36:4), pp. 537–563. (https://doi.org/10.1177/0170840614559257).

Moogk, D. R. 2012. "Minimum Viable Product and the Importance of Experimentation in Technology Startups," *Technology Innovation Management Review*, pp. 23–26.

Nielsen, J. 1989. "Usability Engineering at a Discount," in Proceedings of the Third International Conference on Human-Computer Interaction on Designing and Using Human-Computer Interfaces and Knowledge Based Systems (2nd Ed.), USA: Elsevier Science Inc., September 1, pp. 394–401.

Nunamaker, J. F., Briggs, R. O., Derrick, D. C., and Schwabe, G. 2015. "The Last Research Mile: Achieving Both Rigor and Relevance in Information Systems Research," *Journal of Management Information Systems* (32:3), pp. 10–47. (https://doi.org/10.1080/07421222.2015.1094961).

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of Management Information Systems* (24:3), pp. 45–77. (https://doi.org/10.2753/MIS0742-1222240302).

Peirce, C. S. 1878. "Illustrations of the Logic of Science: How to Make Our Ideas Clear," *The Popular Science Monthly* (XII), (E. L. Youmans and W. J. Youmans, eds.), pp. 286–302.

Poon, C. S. K., Koehler, D. J., and Buehler, R. 2014. "On the Psychology of Self-Prediction: Consideration of Situational Barriers to Intended Actions," *Judgment and Decision Making* (9:3), p. 19.

Prat, N., Comyn-Wattiau, I., and Akoka, J. 2015. "A Taxonomy of Evaluation Methods for Information Systems Artifacts," *Journal of Management Information Systems* (32:3), pp. 229–267. (https://doi.org/10.1080/07421222.2015.1099390).

Pries-Heje, J., Baskerville, R., and Venable, J. R. 2008. "Strategies for Design Science Research Evaluation," in *ECIS 2008 Proceedings*.

Rhaiem, K., and Amara, N. 2021. "Learning from Innovation Failures: A Systematic Review of the Literature and Research Agenda," *Review of Managerial Science* (15:2), pp. 189–234. (https://doi.org/10.1007/s11846-019-00339-2).

Rogers, Y., Sharp, H., and Preece, J. 2011. *Interaction Design: Beyond Human-Computer Interaction (3rd Ed)*, Chichester: Wiley.

Rosson, M. B., and Carroll, J. M. 2002. "Scenario-Based Usability Engineering," in *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, London, England: ACM, pp. 413–413. (https://doi.org/10.1145/778712.778776).

Savoia, A. 2011. *Pretotype It*, (2nd ed.). (https://www.pretotyping.org/uploads/1/4/0/9/14099067/pretotype_it_2nd_pretotype_edition-2.pdf).

Savoia, A. 2019. The Right It: Why So Many Ideas Fail and How to Make Sure Yours Succeed, HarperOne.

Serrat, O. 2017. "Design Thinking," in *Knowledge Solutions: Tools, Methods, and Approaches to Drive Organizational Performance*, O. Serrat (ed.), Singapore: Springer, pp. 129–134. (https://doi.org/10.1007/978-981-10-0983-9_18).

Shahin, M., Ali Babar, M., and Zhu, L. 2017. "Continuous Integration, Delivery and Deployment: A Systematic Review on Approaches, Tools, Challenges and Practices," *IEEE Access* (5), pp. 3909–3943. (https://doi.org/10.1109/ACCESS.2017.2685629).

Silverman, D. 1998. "Qualitative Research: Meanings or Practices?," *Information Systems Journal* (8:1), Wiley Online Library, pp. 3–20.

Sonnenberg, C., and vom Brocke, J. 2012. "Evaluations in the Science of the Artificial – Reconsidering the Build-Evaluate Pattern in Design Science Research," in *DESRIST 2012 - Design Science Research in Information Systems. Advances in Theory and Practice*, pp. 381–397. (https://doi.org/10.1007/978-3-642-29863-9_28).

Tremblay, M. C., Hevner, A. R., and Berndt, D. J. 2010. "Focus Groups for Artifact Refinement and Evaluation in Design Research," *Communications of the Association for Information Systems* (26). (https://doi.org/10.17705/1CAIS.02627).

Venable, J. 2006. "A Framework for Design Science Research Activities," in Emerging Trends and Challenges in Information Technology Management: Proceedings of the 2006 Information Resource Management Association Conference, Idea Group Publishing, pp. 184–187. (https://espace.curtin.edu.au/handle/20.500.11937/24049).

Venable, J., Pries-Heje, J., and Baskerville, R. 2012. "A Comprehensive Framework for Evaluation in Design Science Research," in *DESRIST 2012 - Design Science Research in Information Systems. Advances in Theory and Practice*, pp. 423–438. (http://link.springer.com/10.1007/978-3-642-29863-9_31).

Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: A Framework for Evaluation in Design Science Research," *European Journal of Information Systems* (25:1), pp. 77–89. (https://doi.org/10.1057/ejis.2014.36).

Virzi, R. A. 1989. "What Can You Learn from a Low-Fidelity Prototype?," *Proceedings of the Human Factors Society Annual Meeting* (33:4), pp. 224–228. (https://doi.org/10.1177/154193128903300405).

Weick, K. E. 1989. "Theory Construction as Disciplined Imagination," *The Academy of Management Review* (14:4), Academy of Management, pp. 516–531. (https://doi.org/10.2307/258556).

Yin, R. K. 2003. *Case Study Research: Design and Methods*, (3rd ed.), Thousand Oaks, CA: Sage.

Zittrain, J. L. 2006. "The Generative Internet," *Harvard Law Review* (119:1974), p. 68.