



HAL
open science

What does proof-of-concept (POC) really do? A systematic comparison of generativity and robustness of POC practices

Caroline Jobin, Pascal Le Masson, Sophie Hooge

► To cite this version:

Caroline Jobin, Pascal Le Masson, Sophie Hooge. What does proof-of-concept (POC) really do? A systematic comparison of generativity and robustness of POC practices. 13th International Workshop on Design Theory of the Special Interest Group of the Design Society, Jan 2020, Paris, France. hal-03079478

HAL Id: hal-03079478

<https://hal.science/hal-03079478v1>

Submitted on 17 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

WHAT DOES PROOF-OF-CONCEPT (POC) REALLY DO? A SYSTEMATIC COMPARISON OF GENERATIVITY AND ROBUSTNESS OF POC PRACTICES

Caroline Jobin^{1,2}, Pascal Le Masson¹, Sophie Hooge¹

¹ MINES ParisTech, PSL University, Centre for management science (CGS), i3 UMR CNRS, 60 Bd St Michel 75006 Paris, France

² les Sismo, 30 rue du Temple 75004 Paris, France

Abstract

Although there has been an increasing use of 'proof-of-concept' in practice, the literature is highly poor on the subject. We investigate a basic research question: what does proof-of-concept really do? We perform an historical study and a case study of PoC practices of a French design studio 'The Sismo'. Generativity and robustness allow us to compare PoC practices related to TRLs, medicine, security, start-up, and design. The study clarifies the polysemous practices to help stakeholders to know why and how PoC is a design tool to establish and sustain collaborative exploratory communities.

1. Introduction

As pointed out by the Richelieu's Committee, a network of French innovative companies, for which "proof of concept has been the subject of special attention for several years", there has been an increasing use of this term in practice over the last years (Comité Richelieu and Médiateur des entreprises, 2019). From start-ups and large groups, to research centres, through public institutions, all these typologies of organisation have embraced this terminology inherited from the Technology Readiness Levels developed at NASA in 1970-1980s (Mankins, 1995; Sadin et al., 1989). Such an enthusiasm has to be considered in the context of contemporary innovation challenges that suppose the ability to explore unknowns ever more complex and vaster (Le Masson et al., 2010). Establishing and sustaining original and specific exploratory communities, including actors with heterogeneous expertise who belong to various institutions and often without previous collaboration experience, is an increasingly studied way to face these challenges (Chesbrough, 2006; Star and Griesemer, 1989). Although practitioners recognise in proof of concept (PoC) some virtues such as being a "flexible object in its relations between economic actors" and "a step towards innovation and value creation, a learning step that is often decisive", they also regret that "insufficiently or poorly apprehended, the POC can result in disappointments for both [parts]" (Comité Richelieu and Médiateur des entreprises, 2019). If we pay attention to what the scientific literature tells us about PoC, we realise that the latter is very poor and quite entirely related to the Technology Readiness Levels (Héder, 2017; Mankins, 2009) which today only represents a small facet of its practical uses. Furthermore, this notion anecdotally appears or is even absent in handbooks where we would expect to find it, for instance, in research on new product development (Golder and Mitra, 2018) or on design thinking (Plattner et al., 2018). Therefore, in view of the practical issues highlighted and the sharp existing gap between literature and practice of PoC, we would like to investigate in this paper a basic question that is: what does proof-of-concept really prove or do? The study aims at clarifying the polysemous practices around PoC and helping stakeholders to know why and how PoC is a design tool to establish and sustain collaborative exploratory communities through PoC. To achieve these contributions, we will proceed in two steps: first, an historical investigation of the different branches that constitute the genealogical tree of PoCs, then a case study of a rich practice of PoC in a design agency called 'the Sismo'. We will compare PoC practices in the light of an analytical framework based on two performance dimensions that are generativity and robustness (Hatchuel et al., 2011). In other words, we will investigate (1) what kind of effects are produced and (2) to what extent are these effects of quality? In particular, we will use the scientific method and more precisely the statistical method that follows a codified and widespread learning model (Hacking, 1983; Radder, 2009) to more precisely define this analytical framework and describe how the Sismo's PoCs diverge from it, follow it or do a little both. The plan of this paper is as follows. First, we will conduct

a quick literature review to position the research and present the analytical framework before introducing our methodology and developing the analytical framework. Then, we will conduct the systematic comparison of PoC practices through two steps: an historical study and a case study of the Sismo's PoCs. Finally, we will enlighten the main findings and return to the research question to discuss the contributions.

2. Literature review

2.1. 'Proof-of-concept' genesis

As far as we have been able to trace back, the roots of 'proof-of-concept' came from the Technology Readiness Level (TRL) scale originally conceived at National Aeronautics and Space Administration (NASA) in 1970-80s (Mankins, 2009). Following President Reagan's encouragement for more cooperation between NASA, and civil and commercial as well as international partners (Reagan, 1988), Sadin and colleagues (1989) proposed a new technology development strategy. They introduced a paradigm shift in strategy (from pull to push strategy); instead of entirely driving technology development in relation to the needs of large space programs such as Apollo, the development of some technologies should be permitted regardless space program. Indeed, a framework was needed to assess "how proven a certain technology was" in a context where "they cannot necessarily prove their feasibility in an actual space mission at the time of development" (Héder, 2017). "Proof-of-Concept demonstrated analytically and/or experimentally" described the third level of one of the first established scale (Sadin et al., 1989).

2.2. PoC as an experimental method

Whatever its context and field of application, PoCs in contemporary practice share a common property: its experimental dimension to provide evidence in order to support an idea. Bacon and Galileo are strong advocates of experimental method such as Bernard (1865) in medicine. Such a practice is today often associated to the scientific method (Hacking, 1983) which experienced growth since the mid-1970s and is related to a "strong reliance in statistical methods" (Radder, 2009).

2.3. Generativity and robustness of PoC

Generativeness (or generativity) and robustness are two performance dimensions that Hatchuel and colleagues (2011) used to compare a sampled of design theories. According to them, both notions "generalize two well known series of design criteria that are traditionally discussed in the literature" (Hatchuel et al., 2017). The word 'generativity' comes from the Latin *generativus*, which means "that causes, that creates, that produces" whereas the word 'robustness' comes from the Latin *robustus*, which means "solid, hard, strong, resistant".

Given the current diversity of uses of PoC, we would like to investigate in this paper the following research question: **what does proof-of-concept really do?**

3. Methodology

Our research approach consisted of two steps: (a) an historical investigation of the different branches that constitute the genealogical tree of PoCs (Technology Readiness Levels, medicine, software security, start-up and design) and (b) a case study of a particular rich practice of PoC in a design agency named 'the Sismo'. For the first step, data collection was ensured in scientific literature and grey literature and enriched by meetings with practitioners conducted in the context of the intervention research (Hatchuel and David, 2008; Radaelli et al., 2012) that the main author carried out at the Sismo since October, 2018 as part of her PhD. For the second step, data was acquired through two complementary channels. On the one hand, an 1h30-interview was conducted with the Sismo's employee who had the most experience in the agency. She was also the project manager for two projects (A. and B.) and became the referent in the agency regarding PoC. Project A. aimed at designing a new governmental employment agency model to foster social dialogue and accelerate the return to work and

Project B. aimed at, in client stores, rethinking the discovery of product and service lines and developing new services to their customers. During the interview, she was asked questions related to generativity and robustness of the Sismo's PoCs. On the other hand, this interview was mediated and enriched through the intervention research conducted at the Sismo. In both steps, a systematic comparison will be performed between PoC practices in the light of an analytical framework based on two performance dimensions that are generativity and robustness. (1) We define generativity as the ability to produce effects that go beyond the initial knowledge of the designer community. (2) We define robustness as the ability to produce "expected performances despite being subjected to uncertainties or hard-to-control disturbances" (Hatchuel et al., 2011). In particular, we will use the scientific method and more precisely the statistical method that follows a codified and widespread learning model (Hacking, 1983, Radder, 2009). Such a model will allow us to more precisely define this analytical framework that will be used to describe how the Sismo's PoCs diverge from it, follow it or do a little both. (1) In terms of generativity, statistical method allows to reject (or not reject) a null hypothesis if the observed significance level (p-value) is less than or equal (more than) to the chosen value of alpha (type I error) (Karlsson, 2016). (2) In terms of robustness, we choose to use the main requirements of scientific method outlined by Sekaran (1992): (1) *Purposiveness*: The researcher/practitioner has been guided by and communicates a purpose for the research, (2) *Rigour*: A strong theoretical base and a sound methodology are needed to collect the appropriate information and to interpret it adequately; that is, to do research in a trusted manner, (3) *Testability*: The researcher/practitioner can reliably infer whether or not the data support conjectures and he or she controlled for what can influence the results, (4) *Replicability*: It is possible to repeat the study exactly. If controlled repetition of replicable studies is performed, conjectures will be neither supported (or discarded) merely by chance, nor due to differences in method, (5) *Precision and confidence*: These refer on the one hand to how close the findings are to 'reality' and, on the other, to the probability that estimations are correct, (6) *Objectivity*: The conclusions are based on all the relevant facts and are not influenced by the researcher/practitioner's subjective values, (7) *Generalizability*: Refers to the applicability of the research findings to settings different from the studied one, and (8) *Parsimony*: The use of small number of variables and relationships among variables to describe and explain a phenomenon generally makes research frameworks more manageable and useful. This framework was presented to the Sismo's project manager and was used to drive the interview.

4. Historical study of PoC

4.1. PoC in Technology Readiness Levels (TRLs)

As far as we have been able to trace back, the roots of 'proof-of-concept' came from the Technology Readiness Levels (TRLs) originally conceived at NASA (Mankins, 2009). NASA was established in 1958 as an independent agency of the United States Federal Government responsible for the civilian space program to catch up with the advance taken by the Soviet Union. Following the observation that "new system developments suffer from cost overruns, schedule delays and the steady erosion of initial performance objectives", the idea of a 'technology readiness review' emerged (Mankins, 2009). In the 1970s, the NASA's Office of Aeronautics and Space Technology department introduced the concept of 'technology readiness levels' (TRLs) "as a systematic tool that enables assessments of the maturity of a particular technology and the consistent comparison of maturity between different types of technology" (Mankins, 2009). According to the inventors of the first TRL scale, S. Sadin and colleagues, such an instrument that we qualified as a 'boundary object' (Star and Griesemer, 1989) was needed because "the differing perceptions of the researchers and the mission planners between the intended and actual proof of readiness was often the cause of an aborted handoff, or technology transfer, of ART [Advanced Supporting Research Technology] to the SRT [Supporting Research Technology] users" (Sadin et al., 1989). Due to several events (terrible accident of Challenger Space Shuttle, Space Exploration Initiative, ...) that NASA encountered in the mid-to-later 1980s, the use of the TRL scale became more widely accepted and used (Mankins, 2009). When the TRL scale spread in the NASA ecosystem, several versions coexisted that was detrimental to its use. In this context, John C. Mankins (1995) then wrote a paper in which he extended the scale to nine stages and proposed expanded descriptions along with examples for each TRL that is now the standard scale (Mankins, 2009). Indeed, this scale has been

progressively adopted by other U.S. agencies (Department of Defense, Department of Energy, ...). In this 9-level TRL scale in which low levels are associated with low technology maturity and vice versa, 'proof-of-concept' occurred within TRL 3 characterisation ("Analytical and experimental critical function and/or characteristic proof-of-concept") and description (Mankins, 1995). Even though 'proof-of-concept' is related to early maturity and thus early development process phases, the claimed goal is the validation of analytical predictions about applications/concepts formulated at TRL 2 through studies and experiments. This scale also inspired international space agencies and their contractors such as the European Space Agency (ESA) that adopted the TRL scale in the mid-2000s. ESA closely followed the NASA definition of TRLs (Mankins, 1995) canonised through the ISO standard 16290 Space systems established in October, 2014. This scale has globally gained popularity in numerous organisations (Mankins, 2009) in high-tech industry (aeronautical, defence, automobile, ...) and become an innovation policy tool in international institutions such as the European Commission (Horizon H2020 program).

Generativity and robustness: The creation of the TRLs and its associated scale echo what Star and Griesemer (1989) call "generating a series of boundary objects which would maximize the autonomy and communication between worlds". The authors support their statement by sharing that "the creation and management of boundary objects is a key process in developing and maintaining coherence across intersecting social worlds". Indeed, presenting the TRL scale as a tool that fits "both for management of its own instrument technology programs, and also for communicating more effectively with technology researchers and organizations inside and outside" (Mankins, 2009) can, in a big part, explain the enthusiasm generated around it and its massive adoption. If at the beginning, TRLs did meet the definition of boundary objects defined by Star and Griesemer (1989), namely "objects [that may be abstract or concrete] which are plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites", its massive adoption by different institutions has inevitably reduced its robustness. Indeed, in order to ensure its suitability, its adoption was accompanied by an adaptation of the language to better fit organisation or technologies, production patterns, or management structures (Marvel et al., 2012). Moreover, as Héder's paper (2017) pointed out, "the concreteness and sophistication of the TRL scale gradually diminished as its usage spread outside its original context (space programs)", in particular because it does not take into account the existence of implicit related assumptions such as the fact that "the Technology Readiness is an analogue to Flight Readiness". Hence, the interpretation of technology maturity levels has been multifaceted and raised a non-alignment issue that was particularly studied in the technology transfer between universities and industry (McAdam et al., 2009; Jensen and Thursby, 2001). Finally, more than differences in appropriation of scale, different uses of this tool could be described. One example is the Government Accounting Office (GAO) case study reporting the first TRL experiments at the Department of Defense. The report showed that projects that had used immature technologies inherited various difficulties, conversely, programs that integrated TRL6-technologies no longer experienced any additional cost. The main message of this report was therefore that by formalising the maturity of technologies on the TRL scale and by refusing to support programs technologies that are too immature (below level 7 in any circumstances, with the preferred levels 8 and 9), DoD will improve the performance of these acquisition programs (Government Accounting Office, 1999; Jean, 2018).

4.2. PoC in medicine

Following the tradition of medicine as an evidence-based practice and responding to the excesses that have marked our history, the pharmaceutical and biomedical industries are today characterised by one of the most regulated environments related to safety and ethics. In that context, PoC became an 'obligatory passage point' (Callon, 1984) in drug and medical device development process following the philosophy of Cooper's stage-gate (1994). Indeed, PoC is a widely used term to describe the Phase IIa of clinical trials and even described "perhaps the key deliverable of exploratory drug development" (Cartwright et al., 2010). During this stage, "the study drug is tested for the first time for its efficacy in patients with the disease or the condition targeted by the medication. These studies may have up to several hundred patients and may last from several months to a few years" (Corr and Williams, 2009). Cartwright et al. (2010) suggested an interesting alternative definition of PoC borrowing an expression from the code of federal regulations: "POC is the earliest point in the drug

development process at which the weight of evidence suggests that it is 'reasonably likely' that the key attributes for success are present and the key causes of failure are absent". In addition of being a milestone in an internal development process to "allow drug developers to make 'Go/No Go' decisions about proceeding with larger, more expensive studies" (Preskorn, 2014), such studies are a "regulatory step" that must sufficiently convince the competent authorities to allow the release of their agreement to pursue the development process". **Generativity and robustness:** The goal of the PoC is not only to "demonstrates that the drug did what it was intended to do, that is, interacted correctly with its molecular target and, in turn, altered the disease", but also "help[s] determine the correct dosage, common short-term side effects and the best regimen to be used in larger clinical trials" (Corr and Williams, 2009). Contrary to later stages of drug development, PoC "typically involv[es] a small number of subjects and more latitude in statistical requirements" (Preskorn, 2014). If we look at the scientific publications in this field and the structures which incubate biomedical technological innovations, we realise that PoC can play a completely different role than the one described above and overlap with the genealogy of the TRL. Thus, most scientific publications using the term PoC, across all disciplines, come from life science and healthcare researchers working in laboratories whose main objective seems to be the prospect of technology transfer. Moreover, for researcher, stating that he or she has done his or her PoC or achieved TRL3 allows him or her to get access to acceleration programs regarding support and/or budget. This dual system of PoC in this field is problematic and has been reported to us by a Technology Transfer Project Manager at SATT Lutech, a French structure that accompanies innovative projects from academic laboratories to the market. He told us that there were often misunderstandings between researchers and large groups looking to buy back licences. Thus, the researcher may consider having established his PoC by validating the mechanism of action on a few human cells while the industrialist thought that the drug had already been tested on several dozen sick patients.

4.3. PoC in (software) security

A genealogy of PoC also comes from the hacker movement, especially in the world of software. The purpose of PoC is to show that there is a security breach that allows a potential intrusion. This person can be benevolent and reports such defect to its manufacturer or even proposes corrective solutions in order to prevent malicious person to enter the same security breach. Such security breach can constitute an entry point for a virus capable of altering the DNA of the software or I.T. system and thus reprogramming it to divert its initial use or crashing the system. Indeed, due to our increasing reliance on I.T. systems and the new black gold that data represents, such attacks or at least attempts may unfortunately be more and more casual. **Generativity and robustness:** As clearly express in English Wikipedia regarding PoC in security, the PoC is a demonstration that "a system may be protected or compromised, without the necessity of building a complete working vehicle for that purpose". One can take the example of Winzapper "which possessed the bare minimum of capabilities needed to selectively remove an item from the Windows Security Log, but it was not optimized in any way".

4.4. PoC in start-up

One of the main reasons why start-ups perform PoC or declare having done is for its strong power of attraction. With such label, entrepreneurs hope to attract and convince stakeholders that the concept is worth investing in. Stakeholders or PoC targets may be potential internal or external partners, potential clients, investors, buyers, etc. Another reason for B2B and B2BC start-ups that explain why they carry out PoC is related to the widespread practice of open innovation in large groups (Chesbrough, 2003) that are increasingly looking at start-ups. As Bpifrance, a French public investment bank that supports French companies by offering financing solutions to support their projects and development points out, many start-ups and large groups enter into collaboration through PoC. According to Bpifrance, on the side of large groups, PoC "allows to assess the capacity of the solution to meet the expectations of the large group and to see whether collaboration between the two stakeholders can be considered on a larger scale". On the side of start-ups, PoC is an opportunity to "generate revenue by acquiring customers [and] to engage quickly with large groups, with few technical constraints, and demonstrate the value of a start-up's services". In addition, PoC is a term that is increasingly found in support structures such as incubators and accelerators that may or may not be corporate, either as an exit or an entry point. Finally,

PoC is no longer restricted to start-ups but is open to the notion of innovative company which could be all kind of organisation. **Generativity and robustness:** On the one hand, as we stress above, PoC has a strong power of attraction, and that is the reason why support structures such as incubators help entrepreneurs to achieve it. Experiment approach as promoted in Lean Startup (Ries, 2011) then often serves as a reference. On the other hand, PoC in its collaborative shape, tends to generate "disappointments for both the person who delivers it and the person who commands it" (Comité Richelieu and Médiateur des entreprises, 2019) and thus does not lay the foundation for a sustainable relationship beyond PoC. According to Bpifrance, some companies are "promoting a reputation of 'serial Pocer'". We can interpret this trend as a (new) purpose of such collaboration that is not intended to provide validation prior deployment (as the start-up expects) but a new exploration strategy.

4.5. PoC in design

PoC in the design world is often associated with the prototype and prototyping stage. Indeed, it is not uncommon to see the term PoC in the form of 'proof-of-concept prototype' (Yu et al., 2018). It can be explained by the fact that the model, the prototype acts as a proof of feasibility through the notion of materiality (Blomkvist, 2014; Kirsh, 2010). In the same vein, Gay and Szostak (2016) declare: "one of the strengths of design thinking lies, in addition, in its ability to visually translate the senses created, especially through the proof of concept". As a result, the notion of PoC is increasingly reflected in the world of design thinking and design sprint. Even if this term is very weakly present or even absent from the original texts (see for instance Brown 2008; 2009) and seems to appear from nowhere, it has a real stake and interest in the contracting between design agencies and their clients. For instance, for 'the Sismo', an independent design agency, PoC has become a service that can be sold to clients through a quote. Therefore, a remuneration is ensured while the final object, in the broad sense of the term, has not yet seen the light of day. **Generativity and robustness:** In terms of robustness, it was usually the client or the sponsor who had to be convinced by the designers. As in a criminal trial in France (where a part of the jury consists in citizens), the clients are themselves the sole judges of "the fullness and sufficiency of the evidence" (Article 353 of the French Code of Criminal Procedure). What we see more and more, especially at the Sismo, the central point is to convince the manager or the management of the client. Indeed, the methods of proof generation and the levels of evidence expected by the client are increasingly demanding. Before going into more detail about the generativity power of the Sismo's PoCs, it is enriching to look at the case of 'Lille 2020, World Design Capital'. For this edition, it is not only about a global event, but the purpose is also to infuse, integrate and transform the metropolis territory through design. The goal of the organising committee is clear: "to make the territory a laboratory for large-scale experimentation [to] imagine and experiment new ways of designing, developing and deciding" through PoCs. While most of the initiatives are still in their infancy, the initial exchanges among committee members are promising. For instance, they stressed that (1) the general public and media character of the event has attracted project promoters of all kinds, (2) PoC has allowed people from diverse institutions to sit together whereas in another context this could have never happen, and (3) for organisations that have already embarked on experimentations, there has been a certain awareness that the transformation to which they aspire has an echo in their own organisation.

5. Case study of PoC

5.1. Introduction of the Sismo

Founded 22 years ago, the Sismo is an independent design studio, renowned for its strategic approaches to support innovation, for French and international large groups, to smaller companies, through public and cultural institutions. Often awarded for its iconic creations, methodological tools and its public exhibitions, the studio has gradually made the service and experience design the point of convergence of its various expertise. The studio is increasingly recognised for its expertise in PoC as evidenced by its involvement in 'Lille 2020, World Design Capital'. PoC describes one step of their project approach which consists of upstream steps (ethnographic study, co-creative workshop and scriptwriting) and downstream steps (valorisation and deployment).

5.2. Generativity power of the Sismo's PoCs

For designer teams: Due to the nature of the subjects of work related to service and experience design, they must deal with "intangible, immaterial things that it is sometimes difficult [for the Sismo's designers] to represent" and to anticipate. Then, the PoC helps to reduce some uncertainties. The interviewed project manager took the example of project A., "we had worked with advisors on a way to separate the flow of arrival [of job seekers] in three: administrative reception, express reception and second level reception, with the hope of decreasing and removing the long tail, but we were not sure at all [that the expected effect was going to happen]". One another objective of the PoC for the Sismo is to gather evidence necessary to convince stakeholders and then ensure massive deployment and adoption. Thus, as the project manager pointed out, it is not always easy to find the right arguments for both the management teams and the field teams. These gathered elements will feed a "deployment booklet where we put [...] photos and verbatim to prove what we were trying". The PoC is also an opportunity to continue the design, as the project manager said, "it's a playground, it's a board game, you put the cards, the pawns, [...] you have the conditions to continue to design, that's why I say it's a design phase [and not just a test]". *For field teams:* As mentioned above, the Sismo often works on things that are immaterial which require "a strong projection capacity of people". Then, the PoC has a role of conviction through an experiential scenario. Thus, as the project manager emphasised, "it is by making them feel that we convinced them". To support this, she talked about the project P.: "There are some advisers for whom we only managed to convince and board them through the PoC, if I take an example, when we said to some of them, you are going to work on a narrower table, they saw only security problems, while in situation they told us "ah but we are closer [with the job seeker] so we speak less loud and so there is less aggression!". Moreover, the PoC is an opportunity for the field teams to have the means and framework to do and dare to do otherwise. As emphasised by the project manager, "it is rare that they [the field team] have the means to do that!", and thus, "it's important to change their outfit to become [this new character]". More than desired, this appropriation is, almost in all cases, necessary because the Sismo are not often experts in the field of the client organisation or subject of work. Indeed, the challenge is to put the right bases or good cards, pawns on the board game to use the expression of the project manager and "after that it is up to [the field teams] to play!". For example, if we take again the example of project A. with the partition of the reception flow in three, "it is to the advisers to define which questions relate to which reception and to determine if express reception is related to an answer that takes 5, 10 or 15 minutes? etc." *For management teams:* As the project manager pointed out, "in all cases, it [the PoC] makes communicate the headquarters and the field where you have two realities". Thus, the PoC allows a certain awareness of management teams by reconnection with the field. However, certain conditions are necessary to generate this effect. Management teams must come to the PoC in the field and adopt a position of observer with empathy or even take the posture of the field teams, and this on a time long enough. For example, in project B., the project manager recounted the client's comments: "She said that it had been such a transformation for her, it was good, ..., it had been a long time since she had not had the opportunity to spend 6 weeks at the same store, [...] which allowed her to realise that sometimes agents were asked very complicated things". The PoC also allows to "give an account that some things need to be improved" and thus to open new (transformation) programs and projects and position these as priorities. Furthermore, just like the field teams, the PoC is an opportunity to dare and test new things. As the project manager commented in the context of project B., "what amused me was that the client [...] was offering iterations herself, she was proposing things that were not always very interesting [...] but we let them do so that they see by themselves that it does not work [...]. I felt that she was in this dynamic of saying there is a problem and I am looking for solutions but that it has its limits ... Fortunately we were there for the mediation and hopefully, they still let us do the restitution because they had made certain shortcuts, taken certain things at face value, they did not have the overall view as we can have".

5.3. Robustness power of the Sismo's PoCs

To begin with, the first statement of the project manager in reaction to the analytical framework of Sekaran (1992) was: "Overall, these criteria well describe robustness". For most of these criteria, we will not only provide quotes from the project manager interview, but also comments to mediate and

complete citations to ensure a more comprehensive view of the Sismo's practice and a finer systematic comparison with the analytical framework. **Purposiveness:** "For the Sismo, it works obviously, [...] behind this there is the hypothesis we want to test". **Rigour:** "There, honestly, we don't take the time to do it well, we rarely have a very extensive theoretical basis, besides, I don't really know where and how to look, I'm bad at doing these things. [...] I have no personal knowledge of ethnology literature." This statement should be moderated; they always start their work with an enrichment phase, but which is more empirical than theoretical. They preferably perform immersions, interviews and basic research on the internet than investigate scientific literature. "At the method level, we don't necessarily have the means to compare in two groups [...] but we try to vary the means to collect the data (interviews, observations, etc.)." Indeed, there are few situations where they will use a control sample to perform an A/B testing analysis. They prefer to mix the way to produce knowledge through different modes of observation on the observed sample. Their sampling does not seek to be random, and this property depends on the level of analysis chosen, i.e., the agency/store was chosen according to certain criteria and was therefore not chosen randomly, on the other hand the people who had frequented these places were random. In addition, people who was selected for further observations or interviews was often chosen because designers had observed unusual behaviours or things that surprised them. **Testability:** "It's complicated, on project B., we don't control the number of people a day, ... it would be distorted the situation to control it, ... and then we did [experiment] it over six weeks". In fact, experiments are not carried out in the laboratory or in the chamber where high confinement can be performed but, in the field, and in real life (or almost) where control is more difficult. Moreover, it is often not clear, *a priori*, what parameters need to be controlled. **Replicability:** "Yes, concretely, we could do PoC in several situations [...] but there are budgetary issues, [...] we had been paid to conduct 3 PoCs [PoC in 3 agencies], we didn't have the means and the time to test everything [all the concepts in each agency], but they [the clients] did [tests] in other agencies." Admittedly, PoC can be replicated, however the observed effects are never only related to the new object alone but to the object/environment pair, so it is important to clarify this interaction so as not to produce abusive inferences. In addition, there is a feedback effect between the observer (the designer) and the observed so the conditions of knowledge production are plural. **Precision and confidence:** "It is clear that we are always trying to be closed to reality, even if we can found few counter-examples [...] we try, during PoC, on the advice of [a middle manager of the Sismo], to be careful not to generalise an observation on the basis of one person." There is a desire in the Sismo's projects to get closer to what they call 'reality', e.g., with tests in the operational environment, for instance, an agency that already hosts jobseekers, an existing store in a shopping centre. However, sometimes the test conditions can be more favourable than the 'reality', for example, by selecting a voluntary agency, supporting the field teams with advice or actions, ...). **Objectivity:** "Of course, with PoC I have concrete evidence, a verbatim, a behaviour, a film, ..., but I think that where there is still a subjective side [...] [it is related to the fact that PoC] is still part of the design phase, we can still remain in the intuition". In view of her statement, the project manager listed the methods of proof that she considered admitted, without asking whether these proofs or their nature were enough for the receiving party, whereas the importance of the way the data was collected, and its related context was mentioned. **Generalizability:** "For Project A., this is not the same issue that the system works everywhere [with reference to project B.], it is a public service, [...] they have the challenge that the reception works in all types of agencies with different audiences [...]. In Project B., they can be more authoritarian, it's their brand, [...] I think they won't test things anywhere else because the sample is representative everywhere [in all their shops]." To complete, each context of observation is quite different but the question to answer behind is whether these differences have a real impact. The notion of representativeness is not something that is directly sought, even though it is often one of the criteria for selecting the field. Knowing that the size of their sample is limited, there is little discussion about statistical inference. **Parsimony:** "When [a colleague] came to me for advice for a PoC [...], she said she wanted to test a lot of things at the same time, I told her to be careful about that."

6. Main findings and discussion

To investigate the basic research question about what proofs-of-concept really do, we systematically compare PoC practices through an historical study and a case study regarding an analytical framework

based on two performance dimensions that are generativity and robustness (Hatchuel et al., 2011). Through the historical review of PoC related to TRLs, medicine, security, start-up and design, we have seen that PoCs are used in different contexts with different objectives and go beyond boundary objects based on a demonstration of feasibility for the purpose of contractualisation. While PoC practices seem to share a common property - federate and structure a collaborative exploration community -, PoCs produce various effects depending on why and how the PoC was designed and carried out. Through the case of the Sismo, we have seen that their PoCs have plural generative dimensions and they are of different natures according to the typology of stakeholders: (1) for designer teams: reduce some uncertainties, gather evidence necessary to convince stakeholders and then ensure massive deployment and adoption, continue the design, (2) for field teams: conviction through an experiential scenario, an opportunity to have the means and framework to do and dare to do otherwise, (3) for management teams: awareness by reconnection with the field, namely the need to open new (transformation) programs and projects, and an opportunity to dare and test new things. These three typologies allow to go beyond the classical client-supplier partition of PoC by opening the way to a more collaborative model. Thus, designer teams often do not reduce themselves to the Sismo's designers and can also be composed of people from the client organisation and civil society. In terms of robustness or guarantee of evidence, a further comparative study between proof of concept and scientific proof based on hypothesis testing should be considered. Indeed, it is commonly accepted that PoC constitutes proof but may not be complete, which raises ambiguity. One can then wonder what natures the constraints are set aside also whether this ambiguity is desired or on the contrary should be better controlled. We hypothesise that the more cautious the designers are in identifying uncontrolled variables, the better the scientific evidence, and the more the PoC will be generative according to these variables.

References

- Bernard, C. (1865), *Introduction à l'étude de la médecine expérimentale*, JB Baillièrre et fils, Paris.
- Callon, M. (1984), "Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay", *The sociological review*, Vol. 32 No. 1_suppl, pp. 196-233. <https://doi.org/10.1111/j.1467-954X.1984.tb00113.x>
- Cartwright, M. E., Cohen, S., Fleishaker, J. C., Madani, S., McLeod, J. F., et al. (2010), "Proof of concept: a PhRMA position paper with recommendations for best practice", *Clinical Pharmacology & Therapeutics*, Vol. 87 No. 3, pp. 278-285. <https://doi.org/10.1038/clpt.2009.286>
- Chesbrough, H. (2006), *Open business models: How to thrive in the new innovation landscape*, Harvard Business Press.
- Comité Richelieu and Médiateur des entreprises (2019), *Osez l'innovation, De l'idée à l'industrialisation : réussissez votre preuve de concept*.
- Cooper, R. G. (1994), "Third-generation new product processes", *Journal of Product Innovation Management*, Vol. 11 No. 1, pp. 3-14. [https://doi.org/10.1016/0737-6782\(94\)90115-5](https://doi.org/10.1016/0737-6782(94)90115-5)
- Corr, P., and Williams, D. (2009), "The pathway from idea to regulatory approval: examples for drug development", In: Lo, B. and Field, M.J (Eds.), *Conflict of Interest in Medical Research Education and Practice*, National Academies Press, Washington, DC, pp. 375-383.
- Government Accounting Office (1999), *Better management of technology development can improve weapon system outcomes, Best practices*, United States General Accounting Office.
- Golder, P. N. and Mitra, D. (Eds.). (2018), *Handbook of research on new product development*, Edward Elgar Publishing. <https://doi.org/10.4337/9781784718152>
- Hacking, I. (1983), *Representing and intervening: Introductory topics in the philosophy of natural science*, Cambridge University Press.
- Hatchuel A. and David A. (2008), "Collaboration for management research", In: Shani, A.B., Mohrman, S.A., Pasmore, W.A., Stymne, B., and Adler, N. (Eds.), *Handbook of Collaborative Management Research*, Sage, pp. 143-162. <https://doi.org/10.4135/9781412976671>
- Hatchuel, A., Le Masson, P., Reich, Y., and Weil, B. (2011), "A systematic approach of design theories using generativeness and robustness", *Proceedings of the 18th International Conference on Engineering Design (ICED 11)*, Lyngby/Copenhagen, Denmark, August 15-19, 2011, *Impacting Society through Engineering Design, Vol. 2: Design Theory and Research Methodology*, pp. 87-97.
- Hatchuel, A., Le Masson, P., Reich, Y., and Subrahmanian, E. (2018), "Design theory: a foundation of a new paradigm for design science and engineering", *Research in Engineering Design*, Vol. 29 No. 1, pp. 5-21. <https://doi.org/10.1007/s00163-017-0275-2>

- Héder, M. (2017), "From NASA to EU: The evolution of the TRL scale in Public Sector Innovation", *The Innovation Journal*, Vol. 22 No. 2, pp. 1-23.
- Jean F. (2016), *L'échange génératif de technologies innovantes : engagement conceptif et conception de la valeur [PhD Thesis]*, PSL Research University.
- Jensen, R., and Thursby, M. (2001), "Proofs and prototypes for sale: The licensing of university inventions", *American Economic Review*, Vol. 91 No. 1, pp. 240-259.
- Karlsson, C. (Ed.) (2016), *Research methods for operations management*, Routledge. <https://doi.org/10.4324/9781315671420>
- Le Masson, P., Weil, B., and Hatchuel, A. (2010), *Strategic management of innovation and design*, Cambridge University Press. <https://doi.org/10.1017/CBO9780511779916>
- Marvel, J. A., Saidi, K., Eastman, R., Hong, T., Cheok, G., and Messina, E. (2012), "Technology readiness levels for randomized bin picking", *In Proceedings of the Workshop on Performance Metrics for Intelligent Systems, College Park, Maryland, March 20-22, 2012*, pp. 109-113.
- Mankins, J. C. (1995), *Technology Readiness Levels: A White Paper*, Advanced Concepts Office, Office of Space Access and Technology, NASA.
- Mankins, J. C. (2009), "Technology readiness assessments: A retrospective", *Acta Astronautica*, Vol. 65 No. 9-10, pp. 1216-1223. <https://doi.org/10.1016/j.actastro.2009.03.058>
- McAdam, R., Miller, K., McAdam, M., and Teague, S. (2012), "The development of University Technology Transfer stakeholder relationships at a regional level: Lessons for the future", *Technovation*, Vol. 32 No. 1, pp. 57-67. <https://doi.org/10.1016/j.technovation.2011.08.001>
- Plattner, H., Meinel, C., & Leifer, L. (Eds.). (2018), *Design Thinking Research: Making Distinctions: Collaboration Versus Cooperation*, Springer. <https://doi.org/10.1007/978-3-319-60967-6>
- Preskorn, S. H. (2014), "The role of proof of concept (POC) studies in drug development using the EVP-6124 POC study as an example", *Journal of Psychiatric Practice*, Vol. 20 No. 1, pp. 59-60. <https://doi.org/10.1097/01.pra.0000442938.61575.c2>
- Radaelli, G., Guerci, M., Cirella, S., and Shani, A. B. (2014), "Intervention research as management research in practice: learning from a case in the fashion design industry", *British Journal of Management*, Vol. 25 No. 2, pp. 335-351. <https://doi.org/10.1111/j.1467-8551.2012.00844.x>
- Radder, H. (2009), "The philosophy of scientific experimentation: a review", *Automated experimentation*, Vol. 1 No. 1, pp. 2. <https://doi.org/10.1186/1759-4499-1-2>
- Reagan, R. (1988), *Presidential Directive on National Space Policy 11*, Washington DC, The White House.
- Ries, E. (2011), *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*, Crown Books.
- Sadin, S. R., Povinelli, F. P., and Rosen, R. (1989), "The NASA technology push towards future space mission systems". *Space and Humanity*, Pergamon, pp. 73-77.
- Sekaran, U. (1992), *Research methods for business*, John Wiley & Sons, New York.
- Star, S. L., and Griesemer, J. R. (1989), "Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39", *Social studies of science*, Vol. 19 No. 3, pp. 387-420.
- Yu, F., Pasinelli, M., and Brem, A. (2018), "Prototyping in theory and in practice: A study of the similarities and differences between engineers and designers", *Creativity and Innovation Management*, Vol. 27 No. 2, pp. 121-132. <https://doi.org/10.1111/caim.12242>