The Gaming of Systemic Innovations
innovating in the railway sector using gaming simulation

van den Hoogen, Jop

DOI
10.4233/uuid:b81bf4a5-95f8-4caf-9167-7825b69a5eab

Publication date
2019

Document Version
Final published version

Citation (APA)
van den Hoogen, J. (2019). *The Gaming of Systemic Innovations: innovating in the railway sector using gaming simulation*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:b81bf4a5-95f8-4caf-9167-7825b69a5eab

The Gaming of Systemic Innovations

innovating in the railway sector using gaming simulation

Jop van den Hoogen

The Gaming of Systemic Innovations - innovating in the railway sector using gaming simulation

Jop van den Hoogen

The Gaming of Systemic Innovations - innovating in the railway sector using gaming simulation

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Monday 7 January 2019 at 15:00 o'clock

by

Jop VAN DEN HOOGEN
Master of Science in Business Administration, Radboud University Nijmegen
born in Arnhem, the Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof. mr. dr. J.A. de Bruin	Delft University of Technology, promotor
Prof. dr. ir. S.A. Meijer	KTH Royal Institute of Technology, copromotor

Independent members:

Prof. dr. ir. A. Verbraeck	Delft University of Technology
Prof. dr. M.A. van der Steen	Erasmus University Rotterdam
Prof. dr. E. Subrahmanian	Carnegie Mellon University, United States
em. Prof. dr. L. de Caluwé	Vrije Universiteit Amsterdam
em. Prof. dr. J.H.G. Klabbers	University of Bergen, Norway
Prof. mr. dr. E.F. ten Heuvelhof	Delft University of Technology, reserve member

ProRail
This research was funded by ProRail B.V. through the Railway Gaming Suite Project.

Printed by: Print Service - Ede
Cover design by: Print Service – Ede
ISBN: 978-94-92679-76-5

Preface and acknowledgements

Similar to systemic innovation processes, the writing of this thesis was a chaotic journey. It was a journey I started to undertake in 2011 and just now could finish, by writing these acknowledgements. It was this chaos that I feel really benefitted my thesis and I wish to thank the section I worked at, for somehow allowing for this creative chaos. The five years I have spent in Delft, at the section Policy, Organization, Law and Gaming of the Delft University of Technology really enriched me, both academically as well as personally. Their style of researching issues of organization and governance was new to me. After a short transition period, coming from a different 'rigorous but less relevant' university, the style proved to be very valuable. Especially the eclecticism and theoretical and methodological freedom I have come to appreciate.

The freedom I enjoyed at the section did not develop spontaneously. For this I am highly thankful to my promotors Hans de Bruijn and Sebastiaan Meijer. They provided me with an environment in which I could explore new ideas, experiment with new methodologies and tentatively work out new conclusions while making sure my exploring did not result in trivial and impractical ideas. The positive cynicism Hans displayed helped me to stay critical of my own work or of others' work I wished to incorporate. To this day, with the increasing prevalence of hypes, fake news, and gurus and their mantras, in academia and in practice, I find this stance highly valuable. Sebastiaan Meijer, as my supervisor, gave me the best combination of freedom and guidance. Somehow he ensured progression while still making sure that my PhD was exploratory. The supervisory style also made me feel that I was fully in charge of my PhD all the while knowing I could always fall back to him for support. In addition I have come to respect the way his ideas have implicitly influenced my thesis and me personally. I remember multiple instances where I thought of coming up with radically new and interesting insights and then immediately realized he somehow mentioned this way earlier in the process to me. Learning is indeed serendipitous but Hans de Bruijn en Sebastiaan Meijer somehow managed to steer my learning curve for the better. That is true process management.

My closest colleague was Julia Lo and I thank her for the collaborations we undertook in writing articles and papers, employing gaming simulations and just thinking up new ideas and concepts. I will remember the many heated debates about what the differences are between system awareness and situation awareness (if I recall correctly). I believe we have actually never solved this issue. Here I would also like to thank other colleagues of the 'Gaming team' Bill and Femke. I hope you can continue the good work Delft is doing on making simulations more interactive and collaborative.

This thesis, and me as a researcher, highly benefited from the collaboration with ProRail. Therefore I wish to thank the organization in general and a few people in particular. Emdzad

Sehic and Jelle van Luipen from the innovation department not only helped shaping the games that this thesis studied, but also functioned as ‘klankbord’ for my research. Their enthusiasm for gaming and railways was infectious.

For moral support I wish to thank my coffee drinking buddies in Delft: Qasim, Emiel, Devender, Arman, Giovanni, Carlos and many more, with whom, in a range of different combinations, I enjoyed coffees and discussions about whether the TV series Homeland realistically depicted some of their home countries or not. It was a welcome daily distraction from the PhD work.

Finally, and maybe most importantly, I would like to say thanks to my girlfriend Katharina. She is the one on which I could fall back unconditionally. In some sense a PhD can be a sort of ‘limbo’ and having a friend in this place may demand patience. Finishing this PhD means we can now really start the serious adult life together. When I initially wrote these acknowledgements before the summer of 2018 I wanted to write “I am eagerly awaiting what lies ahead for us.” Happy to say that it is funny how fast things can go..

Bunnik, 8th of December 2018
Jop van den Hoogen

Table of Content

1 Introduction

After years of burdening engineering works around the Netherlands' most crucial node of its domestic railway network, the new Utrecht Central Station was finally commissioned in late 2016. To the outside eye the improvements might seem merely esthetic, as the new passenger terminal looks modern and vast compared to its predecessor. Additionally, passengers might notice the platform changes and the accompanying timetable changes premiering the last month of that year. One might believe such changes are merely incremental, showing a minor improvement over what passengers grown used to. Since the new infrastructural layout is commissioned in 2016 and the new railway station is in use, train passengers might see a new platform (the 8th one, next to the already existing 7 platforms), slightly new departure times, or different platform assignments for their usual trains.

1.1 The underlying systemic changes

However, underlying these seemingly minor changes were years of more systemic change of which the new working of Utrecht Central Station is the first visible symptom. The way railway tracks, switches, overhead wiring and platform assignments now constitute the Dutch system's central node is a result of a radically different design philosophy. Its origin, at least as it became visible over the course of the research underlying this thesis, could be pinpointed to the sector's visit to Japan in 1997. During this visit representatives of the then monopolist train operating company Nederlandse Spoorwegen (NS) saw how, by applying simplicity into the design of infrastructure and timetables, Japan was able to accommodate far more trains per hour with almost the same technical assets. In addition, the Japanese railway system proved to be far more punctual than its Dutch counterpart. With pressures from the Dutch government to cut capital investments and operating costs while accommodating an expected traffic volume increase the lessons to be learned from the Japanese seemed promising.

Being a networked infrastructure, railway systems are inherently inert and hostile to radical innovations (Markard & Truffer, 2006). The functionality of such systems usually is the result of decades of co-adaptation between system elements, so that their specific alignment makes sure that the system performs reliably on a 24/7 basis. Radically new configurations might be more optimal, but getting there is a serious complex task, which runs the risk of breaking this precious alignment. Still, under the high pressures the Dutch railway system was operating, it seemed to be able to apply such radically different design principles during the redesign of its most crucial station. Especially in a multi-actor environment where responsibilities are dispersed over vast numbers of organizational entities, each with their own incentives, instruments and capabilities, the resulting systemic change provides the scholar of systemic innovations with an interesting case study.

A modern day railway system such as the railway system in the Netherlands comprises of railway tracks, signaling systems, railway switches, trains and many more aspects. As such, the system can be seen as a technical system. However, for the working of a railway system, mere technical equipment does not suffice. Traffic dispatchers, train engineers, maintenance personnel are needed to make trains run on the network and institutions are instigated to organize the work in a safe and economically efficient manner. In a railway system therefore many different systems with differing 'openness' jointly work together to serve a common purpose. Technical artifacts, humans and social organization are three elements found in railway systems and studying these systems therefore has to take into account the different rationalities these subsystems have. Railway systems therefore are seen as large socio-technical systems (Geels, 2004; Markard & Truffer, 2006; Wilson et al., 2007; Künneke et al., 2010) as it is a "purposeful system that is open to influences from, and in turn influences, the environment (technical, social, economic, demographic, political, legal, etc.); the people within it must collaborate to make it work properly; and success in implementation of change and in its operation depends upon as near as possible jointly optimizing its technical, social, and economic factors" (Wilson et al., 2007: 102).

The fact that a failure of such sociotechnical system causes more damage to society than the direct value of service not provided, shows that infrastructures are the backbone of the economy (Bouwmans et al., 2006): they are embedded in a countries' economical, social and cultural infrastructure. Therefore, in their functioning, these systems are influenced by both market power and government regulation as its functioning has an impact on private and public interests (Koppenjan & Groenewegen, 2005). Besides direct involved parties such as ProRail and the train operating companies, other parties play a role in the functioning and transformation of railway systems. Organizations like unions and passenger representative bodies and regulatory bodies and inspectorates have an influence on how the railway system is and should function.

In railway systems as sociotechnical systems we are able to distinguish two different forms of complexity: social complexity and technical complexity (De Bruijn & Herder, 2007; Herder et al., 2008; Weijnen & Bouwmans, 2007). Social complexity arises because in sociotechnical systems many actors are involved that are interdependent but have different interests and perceptions (De Bruijn & Herder, 2007). These interdependencies make it hard for actors to predict the consequences of their actions (Koppenjan & Klijn, 2007). The railway system, compared to traditional socio-technical systems in the manufacturing industry (think of technologies, people and institutions together serving to manufacture soap), have more layers of complexity as "events, operations, people, and technical systems are widely distributed in time and space" and "the distributed system is spread across regional, national, and cultural boundaries, leading to additional problems of interoperability" (Wilson et al., 2007: 102). Furthermore within the operational technical system itself, the system which we can see, large parts of the functions that make up the system are performed by human actors in the role of train engineers, operators, ticketing officers and maintenance personnel. This inclusion of human actors adds to the complexity in two ways. Firstly, the control designers of such a system have on the behavior of these human elements is limited. Secondly, these human elements are not designed to be part of the system, only their function is. As these human elements are selected from a larger sample, the population, their goals might not correspond with the overall goals of the system. For instance, a goal of a train conductor might be to have a pleasant working day, which includes letting late coming passenger board the train to ensure a good atmosphere. However in the light of preventing knock-on delays, waiting for late coming passengers might not be beneficial to the performance of the overall system.

Systemic change in such systems is therefore complex due to the interplay between technical and social complexity and the limited influence single stakeholders can have on such change processes. This thesis explores this systemic change in the Dutch railway sector and intends to uncover the constructive and disruptive mechanisms at work during such a process. By doing so, it will study how given the inertia of the sector and its technological artifact it is managing, inertia that is also functional in some regards, the Dutch sector was still able to systemically change its railway system. In addition, this thesis narrows its focus on the role of gaming simulation as a support method to influence the relevant mechanisms at work. Systemic innovation processes, as we will show during the remainder of this thesis, have a logic and dynamic of their own and this provides the involved stakeholder with little direct opportunity to effectuate any lasting influence on these processes. We will posit that the use of gaming simulation, tools that allow for collaborative experimentation with sociotechnical models of reality, are suitable instruments to manage, control or steer otherwise uncontrollable processes. This application of gaming simulation is relatively new and yet to be fully understood. Recent developments in the gaming simulation practice has seen an increased use of the tool to support policy-making and learning processes (Mayer, 2009) but its use as an applied research tool in the context of systemic innovation is a worthy addition to the functionality of gaming.

This first chapter further introduces the concepts of systemic change and gaming simulation, beginning with a brief introduction of the systemic change that the Dutch railway sector saw in recent years. It will also show how the two fields of innovation and gaming simulation, the first from the spheres of the analytical sciences, the latter from the spheres of the design sciences need each other's input for a valuable addition to the current literature on systemic change and gaming simulation. Literature on sociotechnical transitions, in which many valuable frameworks and notions can be found to study systemic innovations, has mostly overlooked the exact practical role of 'the experiment', which a gaming simulation is, and provides little guidelines to set up and execute such experiments and embed these in ongoing processes. On the other hand literature on gaming simulation in innovation processes is scarce. Work that is done on change processes mostly involves organizational change processes and the gaming simulation literature has neglected the context-of-use when such tools are applied to support innovation processes. Through this juxtaposition we arrive at two problem statements and a central research question. This chapter will end with a thesis outline, showing how the subsequent chapter will help in answering the research question in the concluding chapter 9.

1.2 From flexibility to robustness

From the late sixties, the Dutch railway system functioned according to a set of paradigms by which organizations, managers, designers and operators together constituted what and how the railway system should perform its vital functions. In a polycentric country, where many urban areas of around the same size are seemingly randomly located throughout the country, ensuring direct point-to-point connections is problematic. With the advent of the car in this era, this feature of a railway network became especially significant. Therefore, the way the Dutch network, through its then state-owned operator Nederlandse Spoorwegen (NS, Dutch Railways), tackled this issue was by designing many direct connections between the most important nodes of the network. The resulting timetable in 1970 was deemed revolutionary and instigated a sudden increase of 40% in trains traversing the network. In essence, the new timetable introduced three new elements: firstly, the timetable was symmetric meaning that on a certain line a train connection would always offer the same halting pattern, transfer options and travel time. Secondly, the timetable was cyclical. On a train station, trains going to the same destination would always depart in a specific pre-fixed interval. From Utrecht to Amsterdam one could take the train at for instance 12h15, 12h45, and 13h15. Although not new to the industry at large, clock-face timetables were a feature of urban rail transport, trams and subways, it was one of the first instances of the application of these design principles to a nationwide heavy-rail network. Thirdly and most importantly, to ensure timely transfers between two corridors, NS made the timetable resemble a multi-nodal hub and spoke network. On a certain central node of the network, the hub, multiple trains would arrive and depart at about the same time. Passengers could then change trains in matter of minutes, rather than having to wait half an hour for their transfer.

Especially the combination of a hub and spoke network with a clock-face timetable meant dense traffic loads at railway stations at specific moments during an hour. At 14h15 one would see tens of trains arriving and departing the station, while 5 minutes later no train was to be seen. To accommodate these peaks, NS had to design their stations accordingly. It resulted in stations with many railway switches to allow traffic controllers to reroute trains in case of small delays. Given that many trains had to approach the station at the same time, these switches were necessary. In essence, the stations we started to see in the Netherlands became like a plate of spaghetti, where every track would be connected to all platforms. In a not already at full capacity utilized network, this posed no disadvantages.

These disadvantages became however apparent in the subsequent decades. Higher capacity utilization, currently among the highest in the world, showed that the complexity of the infrastructure around stations also affected the speed by which the station could be approached, the difficulties of maintaining acceptable punctuality levels under adverse circumstance and especially the problematic resilience of the network. With so many possible rerouting options, it became cognitively impossible for traffic controllers to decide what option would be best if train operations had been shutdown and had to be restarted. Recent winters, especially those in 2009 and 2010, showed how vulnerable a heavily utilized network is to small-scale delays around central nodes. Parallel to the inherent disadvantages of a hub and spoke timetable, the railway sector expected diminishing public financing of this mode of transport and increased passenger traffic volumes. According to studies between 2005 and 2010, the amount of passengers using the train network would increase by 50% in 2020. Problematic however was that without financial room to massively expand the network the industry had to find other ways to accommodate growing traffic volumes.

The key change, at least according to the proponents of the change, that would transform the Dutch system to one able to accommodate higher volumes against lower costs would be the focus on robustness rather than flexibility. In 1997 NS already found that the way the Japanese railway sector had evolved to its current makeup led to the railway system being able to accommodate higher traffic volumes at lower operational costs while remaining more punctual. As some instigators of the change in the Netherlands, a change that resulted in the specific way Utrecht was eventually overhauled, posited, the difference between Japan and the Netherlands was the Japanese focus on robustness and the high interdependence between timetable design and infrastructure layout. Instead of having spaghetti-like railway station infrastructure layouts, the Japanese system was one with strictly separated corridors, fewer railway switches and more optimal placing of safety signaling along the tracks. In addition, a higher focus on preventive maintenance did away with the need to have many costly rerouting options since trains and tracks would break down less often. In 2016 the overhaul of Utrecht central station showed exactly that: a more lean design approach to the use of rerouting options, less railway switches, more optimal placing of signaling and the subsequent ability to traverse this part of the network with higher speeds. The separation of corridors would also mean that if one train were delayed, it would not cause so-called knock-on delays for other trains. The much used Dutch, but hard to translate term ‘oil-spill working’, where small delays could spread throughout the entire network, would hopefully be a thing of the past.

These changes run parallel to drastic institutional changes to the way the railway system is governed. The official separation between exploitation of train services by NS and the management of the railway infrastructure by ProRail in 2003 left the latter organization with a smaller set of tasks but also with more operational freedom to execute these tasks. In addition, operational performance became more transparent. Under the umbrella company, layers of higher echelons shielded the infrastructure management departments from public scrutiny. Now, as an independent organization, ProRail's performance was under scrutiny of the public and governmental agencies. On top of that, because the train operating company (TOC), now a semi-private company, was contracted on a 10 years franchise basis, ProRail would have to be the safeguard of long term interest of the railway sector in the Netherlands. Hence, from a internal department mostly responsible for executing tasks ordered by higher echelons, now the organization had to decide on what tasks it would do and how. The implementation of the so-called Japanese principles, the principles used to redesign Utrecht Central Station, was eventually taken on by this organization.

1.3 Systemic change

Using principles that are the result of evolutionary processes elsewhere, in this case Japan, is cumbersome and poses a radical departure of what is considered standard practice in the Dutch case. We call such a departure a systemic innovation, which we see as a specific type of innovation. Innovation entails the invention and implementation of something new (Van de Ven, 1986) As such, it is both about the process of creativity, i.e. invention, and the process of bringing creative ideas into being, i.e. implementing them. Systemic innovations are needed when the functionality of the system cannot be guaranteed if the system incrementally changes and leaves its history in tact. Transitions then shift the system towards a new path, a path that promises a better match between what the system offers and what the environment, e.g. passengers and governing bodies, demands. In this case, some within the sector envisioned that leaving the path of flexibility and transitioning towards a path of robustness would increase the overall performance of the Dutch railway system.

Systemic innovations are new configurations of a system in which multiple elements of the system are changed simultaneously and that reorient the developmental path of a system. This definition is based on complexity perspectives on sociotechnical transitions (Alkemade et al., 2009). These innovations not solely tackle (and improve) one specific element of the system (termed incremental) but rather intend to change an entire constellation of system elements. Systemic innovations may also be named architectural innovations (Taylor & Levitt, 2004) in that it differs from radical innovations: it reinforces the core concept that constitutes the product instead of radically changing it (see Henderson & Clark, 1990). In this thesis systemic innovations are also characterized as transitions. They move a system from one technological trajectory, or developmental path, towards another. Such 'other trajectory' might have been plausible as an autonomous way forward in the history of the evolution of the system. However historic decisions have steered the system along another trajectory.

An example of a systemic innovation would be the transition from Qwerty to Azerty computer keyboards, both once viable alternatives to the layout of a keyboard. However a current transition from one to the other would need additional measures in user training, software adaptation, etc. Systemic innovations are collections of changes that individually would have no value (solely changing Qwerty would decrease the 'value' of computers) and as such do not follow automatically from the historical developmental path of the system. Revolutionary change would be a good synonym as it involves breaking down the old system and building, from scratch, a new and potentially better system, with the same building blocks. However, we would like to point out that revolution in itself is not a good thing. Revolution involves replacement of an old system with a new system and this inherently entails three problems: firstly, the replacement might become partial, where the combination of the old and new performs less well than either the old or the new separately. Secondly, how good the new system is, is purely subjective. What one might consider an improvement might be considered deterioration by someone else. Thirdly, in the transition process from the old to the new, the system might deteriorate to such an extent that the business case for the new becomes negative.

A transition towards a railway system designed with Japanese principles in mind will involve changing the infrastructural layout, the way safety signaling is designed, timetables, operator roles and rules, institutional setting of the sector, and even the way passengers use the system. Such systemic innovations are complex but also crucial for systems to remain legitimate if incremental innovations no longer suffice. However, while we as society demand networked infrastructures, such as the railways, to adapt to changing environmental demands, they must carefully balance meeting these demands with activities focusing on providing current functions reliably on a daily basis. All too disruptive change might very well cause the reliability of its core functions to worsen. For example, if a railway infrastructure manager would find that an increased width between railway tracks would improve the overall performance of the railway system, it faces the dilemma to suspend operations for years and provide this improvement or to opt out and maintain the status quo. A common dilemma that many organizations face (March, 1991). However, this dilemma is much more prominent for these types of industries than they are for more 'traditional' product-oriented industries on which literature has mainly focused. Whereas the production and functioning of a product are separated in space and time, services that networked infrastructures deliver are produced and consumed at the same time. It is therefore interesting to zoom in on innovation processes in these 'system-oriented' industries and how they balance exploitation and exploration and to what dynamics this leads in the innovation processes that sector stakeholders undertake.

This thesis zooms in on this dilemma, studying how organizations involved in managing, operating and designing railway systems trade-off so-called exploitative and explorative activities in a multi-stakeholder environment. The overall problem, with which this part of the thesis starts, is the fact that more systemic innovations are hard to implement. The complexities of such an innovation process involve two kinds of coordination problems: how to coordinate separate innovation activities in a multi-actor environment? And how to coordinate innovation activities and already existing exploitative activities? Hence we define the problem statement regarding systemic change in railway system as follows:

"Railway systems will see an increasing need for more systemic innovations, however as systems that are exploited on a continuous basis they are hostile towards these radical changes. In an environment where no central authority exists, the implementation will be highly organic and less controllable. However, systemic innovations are those innovations that mostly, but also potentially, show their value through the synergetic interplay of different smaller scale changes. The needed coherent implementation of all intended changes is, given its highly complex nature, problematic"

The newness of the phenomenon of systemic innovation to the railway sector is what adds to the problem. While more status quo oriented activities of organizations in the railway sector have had time to fully develop, think of asset management and timetable design, more innovative activities are new to the sector (Markard, 2006) and little to no work has been done on describing and improving these activities. It therefore is highly relevant for both academia and society to see how such innovation processes look like and what can be learned for future innovative endeavors in railway systems and similar network-type industries.

1.4 Knowledge gap

Recently we see a rising prominence in research on the transformations of networked infrastructures (Verbong and Geels, 2007; Frantzeskaki and Loorbach, 2010; Lovell, 2012; Markard and Truffer, 2006; Markard, 2011). However, most of this research has focused on the impact of liberalization as means to reconfigure the working of the industry. Given the dramatic changes in recent years to how networked infrastructures are governed in Europa such research topics are important to address. However, the central purpose of innovation is that it allows the infrastructure system to adapt to changing societal demands. This adaptation can be done through better performing the functions, or performing new functions, that society appreciates. New institutional setups will not do this per se, rather they only state the rules of the game by which actors inside our outside the sector might envision, nurture, experiment with, implement and block promising new avenues for improvement. The broader society is in the end not directly helped with a new institutional structure, but with a value proposition, through its sociotechnical artifact, by the complete sector that is in line with the specific demographics, travel patterns, energy use and values of society at that specific time.

In the theoretical frameworks often applied to studying transitions of sociotechnical systems scholars place much emphasis on the experiment. Given the anatomy of the change processes this stream of literature studies, their systemic nature and the scale and scope of the change, we assume such frameworks to be highly valuable for our analysis of systemic innovations. Frameworks from Transition Management (Rotmans et al., 2001), MLP (Geels, 2002), Strategic Niche Management (Kemp et al., 1998; Hoogma, 2002) and Technological Innovation Systems (Carlsson and Stankiewicz, 1991) literature all have the experiment in the center of their theoretical framework. In these experiments, such as R&D laboratories, green field tests, simulation environments or niches, so-called change-inclined regime players (Rotmans and Loorbach, 2009) can explore path dependence defying innovations in a setting where institutional pressures to remain incremental are absent. It is however noteworthy that little to no work exists on how to precisely conduct these experiments. We believe that there are two reasons for this gap in knowledge on transitioning networked infrastructures. Firstly, much of the empirical work involves long-term transitions where the analysis usually occurs a posteriori. It is then cumbersome to trace back which experiments have been conducted where and to retrieve the then used rationale for this experiment and its design. Secondly, the perspectives used to study networked infrastructures have a distinct structuralist flavor and some of these frameworks have been criticized for neglecting the role of agency in change processes (Genus and Cole, 2008). In their analyses on change processes, many scholars prioritize structures and institutions over individual action to explain the dynamics they encounter. However, the local experiment is exactly an instance where local action occurs to move away from the forcing pressures of applicable structures and institutions. This form of agency might well be studied using existing perspectives, but we posit that historically this phenomenon deserved less empirical attention. Such study would then also provide more normative notions on how to conduct experiments, with whom and using what method.

By narrowing down our analysis to the use of one form of experiment, an experiment that we as researchers are involved in designing and employing in real life in the railway sector, we address these two aforementioned shortcomings in the current scientific literature on systemic innovation processes in networked infrastructures. Firstly, as designers of such experiments we are not loyal to any theoretical framework and hence can shift between structuralist perspectives and more agency-focused perspectives as we see fit. Secondly, by being involved in the process, as action researchers, we can observe the local rationale for designing and employing the experiment at the time, in its specific context and given the expected outcomes by involved stakeholders. Thirdly, and most importantly, we could focus on a single kind of experiment within a context-of-use of systemic innovation processes we were able to observe simultaneously. Therefore we were better able to discern the precise role of this experiment in actually contributing to, or disrupting, systemic change.

1.5 Games as experiments

ProRail, the Dutch infrastructure manager, started to use gaming simulation as a tool for experimentation in 2009. Together with the Delft University of Technology, the organization set out to explore the possibilities of employing gaming simulation as a tool to support innovation processes in the railway sector. Titled the ‘Railway Gaming Suite’, the program intended to introduce gaming simulation to the railway sector as means to test innovations in a safe environment. The programmatic nature of the suite could be found in the organization's intention to make gaming simulation a fixed tool. As such, the suite strived for the use of gaming simulation not merely as a one-off ad hoc tool, but rather as a method that innovation managers, project managers and other decision makers could constantly use. Retrospectively, given the current widespread use of gaming and other participatory variants of simulation within the Dutch railway sector, this has been achieved.

Broadly speaking, three different functions of games can be distinguished (Peters et al., 1998): games for learning, games for policy and games for research. Games for learning and games for policy are already well documented. In short these games make a simplified model of reality, enable participants to learn from or teach about this model and translate findings or knowledge back to reality (Peters et al., 1998). For games for learning, their value lies in the fact that games enable students to experience a system rather than learn about them through the use of language. The transmitting of 'gestalt', our mental picture of reality, from teacher to student through the use of language poses problems since language cannot contain all aspects of a 'gestalt' (Rhyne, 1996). The relative merit of games versus other methods has been tested (Peters et al., 1998), for instance learning about complex systems (Ryan, 2000; Bekebrede, & Mayer, 2006), instructing levee patrollers (Harteveld, 2011) and teaching safety instructions in the oil and gas industry (Meijer & Poelman, 2011).

For games for policy, or policy games, the value is its unique functionality: they allow players to try out new roles and allow for exploration of potential and future reality (Ryan, 2000). Furthermore alternative solutions can be tested in a safe environment (Joldersma & Geurts, 1998; Kriz, 2003). However, the solutions and the model that represents this environment are of a higher level of abstraction (Meijer, 2015). Rather than testing out hypotheses, policy games offer the chance create consensus between decision makers through the multilogue mode of communication where people with different perspectives engage with each other simultaneously (Duke, 2011). Outcomes of games therefore not provide decision makers with ready to use decisions; rather games help to create a future memory (Wenzler & Chartier, 1999). The merit of playing together in a simulated environment can be categorized using the five C's framework of Duke and Geurts (2004). Policy games help in understanding 'Complexity', they improve 'Communication', stimulate 'Creativity', and create 'Consensus' and 'Commitment to action'. Research has shown its positive effect on creating complex system awareness providing decision makers with a common language (Joldersma & Geurts, 1998), fostering idea generation in innovation processes (Duin et al., 2009) and increasing organizational learning (Klabbers; 1993; De Caluwé, 1997; Wenzler & Chartier, 1999). Emperical work has shown how games helped in policy interventions such as organizational development programs (De Caluwé, 1997; Ruohomäki, 2003) and business process redesigns (Tsuchiya, 1998; Ruohomäki, 2003).

The third function of gaming simulations is to provide decision makers with a simulated environment to test out hypothesis about the real world. This function of gaming is still relatively uncommon (Meijer, 2012). Where in games for learning and games for policy the knowledge flows from the game to the participant and from participant to participant respectively, in games for research the dominant flow of knowledge is from the game to the observer. This observer, as the name says, does not participate in the game. This is not to say that games for research cannot perform the functions that games for learning or games for policy have, however its main function is to test out hypotheses in a controlled and simulated environment.

Two features of gaming simulation make it especially suitable as an experimentation tool for the railway sector. Firstly, the simulation part allows stakeholders to study railway systems holistically and dynamically, rather than in a more usual reductionist and static manner. Hence, such stakeholders are able to observe and analyze the nonlinear feedback processes and emergent behavior inherent to complex systems. Rather then providing snapshots of the system, gaming simulation allows stakeholders to see the dynamics of a railway system in action. Secondly, the gaming part means that human players are added to the simulation, leading to potentially increased validity if human behavior is an intrinsic part of the system to be studied and its 'rationale' is still poorly understood and hard to capture in computer algorithms. These two important features of gaming simulation allow the organization to observe and experiment with a system that is both complex and sociotechnical, which a railway system is. The behavior of the railway system, and the performances that the system should show, are partially determined by human behavior. Think of traffic controllers, timetable planners, station staff and train drivers. When a sector wishes to optimize such a sociotechnical system, it needs to take into account both the technical and the social parts of the system simultaneously. In addition to these features, gaming's ability to foster creativity, consensus and communication between multiple stakeholders (as mentioned in the works of Duke and Geurts, 2004 on games for policy-making) might be valuable functionalities during a systemic innovation process undertaken in a multi-actor environment.

Although an interesting premise, *gaming simulation can support systemic innovation processes in the railway sector*, little to no research exists on this specific application of gaming simulation. Most of the academic work on gaming and gaming simulation focuses on policy-making, training, and education and hence assumes that games tend to have effects more easily measurable after the employment of the tool. Enhanced cooperation, increased knowledge, and a better awareness of the complexity of policy-making are some of the more usual goals of gaming simulation. These goals are easy to assess, albeit for many methodological issues, through pre- and posttests using surveys or observations. Gaming simulation applied for research and experimentation is different in that such an assessment is almost impossible. Is it all about valid causal claims, making gaming simulation a simple analytical science instrument? Or is the effectiveness of a gaming simulation more than just providing valid causal claims? Klabbers (2009) and Meijer (2009) already stipulated the two core approaches to gaming simulation, from either an analytical science perspective or a design science perspective. Figure 1.1 shows a model of these two perspectives on gaming simulation:

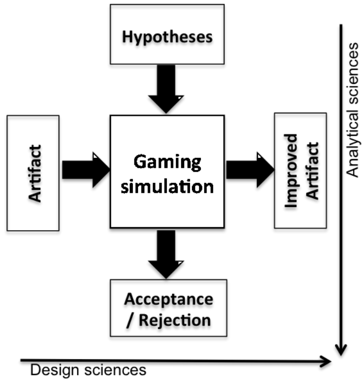


Figure 1.1 Gaming simulation in the analytical sciences and the design sciences, roughly based on Meijer (2009) and Klabbers (2009)

We posit here that we assume innovation processes, as the context-of-use of gaming simulation, can be perceived as both a cognitive-analytical endeavor as well as a constructivist-design endeavor. Hence, one can expect tensions between focusing the design of gaming simulation from either of two perspectives separately. Meijer (2015) acknowledges this tensions and posited that further research is needed for gaming simulation scholars and innovation and system design practitioners to better cope with this tension.

In addition, such tension is not to say overlap may not exist. Designing a game as if it were a classical experiment does not automatically mean it loses its functionality as a game for learning or a game for policy making. Such a game might have functionalities beyond the mere provision of knowledge on causal mechanisms in the system under study: it might create consensus or bring about a dialogue between otherwise separated stakeholders. To take into account this notion of the duality of purpose we call these gaming simulations games for research and design. Next to the fact that we not yet have found one performance indicator (that which makes a game for research and design good) we also have yet to find a sound methodology to assess the extent to which a game performs on this indicator. Hence the second problem statement that this thesis wishes to address:

"Gaming simulation for research and design might be able to influence systemic innovation processes in the railway sector. However, little knowledge exists on both the specifics of applying the method as well as the peculiarities of the process it is intended to influence."

1.6 Central research question

Given the problem statement we need to address two topics that have not been fully understood in conjunction: the design of gaming simulation for innovation, and the process of systemic innovation in networked infrastructures such as the railways. This combination is useful since the improvement of gaming simulation needs the input of the context in which it is used. As Klabbers (2003, 2006) has put it: game design can only be properly done when its context-in-use is fully understood. The design-in-the small (DIS) of gaming simulation is closely linked to the design-in-the-large (DIL) of innovation stakeholders actually bringing about change to the technical and socio-institutional makeup of the railway sector. Given the theoretical gaps we have pointed out, there is need to both understand systemic innovation processes and the role of games as well as the actual design of gaming simulation.

Innovation perspective

To understand the context-of-use in which gaming simulation is employed in the railway sector this thesis adopts a multitude of theoretical frameworks. This is because this context involves technical change (the actual design of an innovation) and socio-institutional change (as people and the norms and rules that are applicable change over the course of an innovation process). In addition, the patterns and mechanisms we wish to uncover are the result of structural forces as well as individual action. Hence no single framework will provide a complete picture of the complexities involved in bringing about sociotechnical change in a multi-actor environment. While in chapter 4 we will discuss the relevant theoretical frameworks, we state here that this thesis will look at the processes under study from an innovation perspective. Hereby we forgo on looking at the process using other lenses, such as ones used in the policy sciences and public administration.

We do this because especially the innovation literature is better equipped to deal two important phenomena we expect to see in systemic innovation processes. Firstly, innovation is a design activity where stakeholders are involved in manipulating an artifact. Therefore the technical change over time is relevant. In addition, this technical change has mutual impacts on the way stakeholders enter or leave the decision arena. Hence, technical change should be endogenous to any model wishing to make sense of systemic innovation processes. We feel that models in the innovation literature therefore are better suited, given its inherent focus on technical change. Secondly, being a design activity where an innovation takes shape over time inherently means that a process consists of a multitude of small decisions and it is this collection of decisions that should be the focal unit of analysis. Usual policy science and public administration frameworks used in the technology and management fields tend to look at one grand decision, such as the dynamics involved in deciding to build an airport or the decisions around implementing a radical innovation, rather than a systemic one. However, in this case we expect that no single grand decision can be distinguished in the first place. Systemic change, we assume, is a long-term process that only after the fact can be seen as systemic and where there is no single moment where an easily demarcated decision arena worked on deciding and implementing 'systemic change'.

Patterns, mechanisms and games
We adopt a process approach towards innovation, studying innovation in action rather than through snapshots. According to Poole and Van de Ven (1989) a good process model always has two complementary components: "The global (macro, long-run) model depicts the overall course of development of an innovation and its influences, while the local (micro, short-run) model depicts the immediate action processes that create short-run developmental patterns" (p. 643). Our descriptive study on systemic innovation processes in the railway sector therefore looks at two different levels of analysis: we look at macro-level patterns that emerge out of the combined effects of underlying mechanisms. This impacts the way we state our research question since we believe that through influencing mechanisms (or micro-level immediate actions), gaming simulation is able to effectuate change to a macro-level pattern. What this relevant pattern is, what underlying micro mechanisms are at work and how games influence these are the starting points for this thesis. Hence, to fill the knowledge gap that is hindering us in addressing the aforementioned problem statements we ask the main research question:

"What mechanisms play a role in driving systemic innovation processes in the Dutch railway sector and in what ways is gaming simulation able to influence relevant macro-level patterns through these mechanisms?"

Both problem statements are highly interrelated and the research question contains elements of both. Whereas the latter problem statement deals with the anatomy of systemic innovation processes regardless of the use of gaming simulation, the first deals with ways gaming simulation can influence these processes. By doing so we embed the practice of gaming simulation in its context-of-use and adhere to the notion of the relation between DIS and DIL of Klabbers (2006; 2009). The specific application of gaming simulation in this thesis, systemic innovation in the railway sector, is new and provides an additional account of how design-in-the-large impacts design-in-the-small and vice versa. Although this thesis takes on a pragmatic approach to the use of a range of theoretical frameworks and research methodologies, the starting point remains that such mutual impact exists.

1.7 Outline and chapter contributions

This thesis tries to bridge three different spheres: theoretical work on innovation processes, the practice of innovation in railway systems, and the methodology of designing and employing gaming simulation experiments. The way these three spheres interact is a structuring force on the research questions we ask and the chapters of this thesis. Furthermore it is in the coalescence of these spheres, which have yet been separated to some extent, where the true contribution of this thesis lies. It does not merely provide a description of innovation processes in the railway sector, it wishes to understand them more deeply using existing theoretical frameworks; it does not look at gaming simulation as an art in isolation, but wishes to place it in existing theoretical frameworks and see its value; it does not merely try to understand gaming simulation, it wishes to see gaming simulation in a broader context of innovation in real life. These goals we have set out deal with a knowledge gap that exists in the current literature on innovation and gaming. As we have mentioned, systemic innovations in networked infrastructures are rare and leave academia with little opportunity for studying them. Hence, case studies on systemic change in these industries are hard to come by. Also for incumbent actors, dealing with these changes, there are little opportunities for learning from similar previous cases. Regarding gaming simulation, the literature so far has mainly focused on games intended to support policy-making and training, and have therefore a presumably more interventionist goal than a game for applied research. By bridging these three separated spheres we wish to contribute to both the practice and the theory of innovation and support game designers in improving the impact their games have on innovation processes. The diagram in Figure 1.2 helps in outlining this thesis and how each chapter supports the subsequent chapter.

Figure 1.2: combining three fields

With this model in Figure 1.2 we show our first and main goal and that is to support the practice of both gaming simulation design and real-life innovation processes by bringing these fields together. This signifies that this thesis seeks to be relevant for policy-makers, game designers, innovators, managers, decision-makers and other stakeholders and helps them in better coping with the complexities that real-world innovation processes encompass.

The relation between innovation and gaming in practice
We start off this thesis by giving a descriptive account on the Dutch railway sector, its history and its evolution over the past 150 years. Given these accounts we provide an overview of the challenges facing the sector as well as ways the sector is currently tackling these challenges. Subsequently Chapter 3 encompasses a grounded theory study into current

incremental innovation processes whereby we look at how current innovation stakeholders give substance to the relation between innovation processes and game design. What models describe and prescribe their actions when innovating and how is then the role of gaming conceived? This analysis helps us in both finding a suitable approach to the design of gaming simulation as well as, as the central tenet of the thesis stipulates, finding suitable theoretical frameworks to analyze the context-of-use of gaming simulation. In this chapter we also compare gaming simulation with computer simulation and life-tests. We base our analysis on a range of life-tests and computer and gaming simulation experiments conducted in real-life in the Dutch railway sector. We therefore first ask the question:

"Given what theoretical frameworks on gaming and innovation processes posit are relevant and valuable features, how do real-life tests, computer simulation and gaming simulation compare on their impact on these relevant features in innovation processes?"

Innovation in theory

Innovation is a process that is not easily demarcated. It is technical, social, and institutional. It is both structured and serendipitous. It can be linear and non-linear. We first seek out to cover a range of existing theoretical frameworks that potentially can help us in analyzing our case studies. Given Van de Ven's call for both describing the macro and the micro, we assume beforehand that no single framework will cover all aspects of the innovation process and therefore we see where which framework is potentially beneficial. Regarding suitable theoretical frameworks we look specifically at how they match with the nature of railway systems as complex and sociotechnical. These theoretical frameworks will also point us to the ontology and epistemology of innovation processes: what are they in essence and how do we know 'innovation processes'? Highly related to the theoretical lens that follows from our study into frameworks, is the methodology by which we use these frameworks and apply this to our research object. Hence we ask two highly related questions in Chapters 4 and 5:

"What theoretical frameworks on change processes in complex and sociotechnical systems best apply to innovating in railway systems?"

"Given the conceptions of innovation processes in the railway sector and the potential role of gaming simulation, what methodology is best suitable to answer the main research question related to both innovation processes and the value of gaming simulation?"

Bridging theory and practice of innovation

This thesis tries to bridge the divide between theory and practice in order for both to gain on the potential synergy. The study presented in this thesis benefited greatly from the fact that the railway system under study, the Dutch national railway system, was undergoing such a systemic change (at least described by many insiders as such) and that we were able to observe the change as it happened. This results in what we believe our first main contribution: that we bring the practice of innovation much closer to the theory of innovation, resulting in empirical lessons for theory and theoretical lessons for practice. In chapter 6 we address the macro-level patterns that emerge when railway systems are innovating. We look at what makes systemic change (high breadth, low depth) unique from more incremental (low breadth, low depth) or radical change (low breadth, high depth) by comparing three cases for each type of innovation. In Chapter 7 we delve deeper in one single case study from the previous chapter and try to find the mechanisms explaining the macro pattern found earlier. We therefore use the following question to structure Chapters 6 and 7:

"What macro-level pattern sets systemic innovation processes in the railway sector apart from other innovation processes and what underlying micro-level mechanisms are at work to create this distinctive pattern?"

Gaming, Innovation, and Real Life
Having uncovered a distinctive macro pattern for systemic innovation processes and the driving micro-level mechanisms underneath it, we look at the value of gaming simulation. This analysis is hence done with an analytical framework distilled from empirical work on actual innovation processes in the railway sector. We contrast this analysis with case studies on gaming using a framework solely derived from theoretical literature on innovation processes.

Whereas this last part of the Chapter 3 results in embedding the practice of gaming in the theoretical work on innovation, the addition of the practice of innovation in Chapters 6 and 7 helps us in fine-tuning the strengths and weaknesses of gaming simulation as an experimental tool in systemic innovation processes. In chapter 8 we provide both a descriptive account of how gaming simulation can influence the mechanisms and patterns found as well as provide concrete normative claims on how to achieve this using different design options for games. Chapter 8 is therefore the crucial chapter and adds to the previous sub question the question related to:

"How does gaming simulation impact the relevant mechanisms that drive systemic innovation processes in the railway sector and what elements of gaming simulation can be manipulated to control this impact?"

For the practitioner: how to design games for innovation
Chapter 8 overlaps both point 5 (the true combination of the three spheres) and point 6 (practical lessons on combining gaming and innovation). In Chapter 9 we build further on the practical lessons of the research finding, mainly by pointing how the games for research and design are all about debriefing, rather than the sole experimental run itself. We provide the gaming practitioner with guidelines on how to use the debriefing in such a way to reap

the benefits of using games as well as cope with the inherent shortcomings of such a tool. Debriefing is seen as the crucial phase of a gaming simulation run for other types of games (Crookall, 2010) but so far little to none is known about debriefing research and design games. In the final chapter we therefore ask the question:

"How, given the uncovered strengths and weakness of gaming simulation in innovation processes in the railway sector, can these strengths and weaknesses be controlled for in the debriefing phase of a gaming simulation exercise?"

2 The Railway Sector: case description and historical context

In this chapter we provide an account of the current situation in the railway sector in the Netherlands. We will discuss its history, the way it is organized and the specific challenges it faces. This chapter is based on widely available information regarding the Dutch railway system. This account will demonstrate the multi-faceted nature of the evolution of the system, over time, and the potential technical and social complexity of reorienting such evolution.

2.1 History of Rail Transport

The initial phase of the development of the railways, or its emergence, was between the end of the 18th century and the beginning of the 19th century, when developments in steam engine technology caused the start of the first railway services in the United Kingdom. Until then, railway transport only served to transport coals within and between mines. For this purpose, the northern parts of England as well as parts of Wales already had an extensive railway infrastructure where carriages were moved using horses or gravity (Freeman & Aldcroft, 1985). After Richard Trevithick adapted the steam engine for use in trains and engineer George Stephenson made the railway tracks suitable for passenger transport. The first two passenger railway services started in northern England: Liverpool and Manchester and Stockton and Darlington were now connected. The railway system was in an acceleration phase: until 1850, the network grew to such an extent that it gave almost complete coverage for the British isle (Freeman & Aldcroft, 1985). In the following 100 years, the railway system stabilized and the network has since then been only slightly expanded. The rising popularity of air and car travel brought about a decline in the usage of the railway system: the system reached the decline phase.

In 1839 the train service from Amsterdam to Haarlem commenced and signaled the beginning of the development of the railway network in the Netherlands. Contrary to popular belief, at least in the Netherlands, the Dutch railway network is not impressively large in size. Not taking into account double and quadruple tracks, the length of the total railway network is around 3000 kilometers. Relative to the size of the total population, the Netherlands therefore has compared to other European countries a relatively small amount of railway track per inhabitant. Furthermore the daily 1.2 million passengers can only board and leave the train at around 300 railway stations. It is this very fact that defines the crux of the problem in the Netherlands. Although traffic volumes are among the top in Europe, the means by which this has to be achieved are far more limited than in other modern countries. Compared to a region similar in size and population, the Nordrhine-Westphalia area in Germany has 4500 kilometers of railway tracks and 900 railway stations at its disposal. Several reasons exist for explaining the underdevelopment of the Dutch railway network.

First of all, the underdevelopment can be historically explained. In the Netherlands, railway transport developed relatively late and only as a response to Belgium's plans to build a railway line between the port of Antwerp and the industrial Ruhr area in Germany (Veenendaal, 2001). The late adoption of railway technology is partly due to the prominence of water transport as the Netherlands has an extensive network of rivers and canals on which at that time transport was cheaper than other modes of transport. When benefits became clear at that time, the government planned a railway link between Amsterdam and Germany. In the meanwhile, private investors planned a railway link between Amsterdam and Rotterdam via The Hague. Both lines were realized around 1850 at which point the UK had already a national railway network. Until the end of the century, the network grow to the size as we know it today, albeit that certain lines were demolished due to lack of demand or due to damages caused by World War II.

A second reason concerns the urban planning dominant in the Netherlands. When after World War II demand for housing was increasing and in the decades afterwards many new villages, towns and suburbs were built, urban planners deemed proximity to a railway line important. Geographically we can see how cities like Helmond, Almere and Amersfoort expanded along the railway tracks that were already there. Most notably Utrechts' expansion follows exactly the trajectories of the northern branch of the railways (Maarssen), western branch (Vleuten, Leidsche Rijn) and southern branch (Houten). Furthermore in urban planning, especially at the end of the previous century and the beginning of the 21st century, commercial centres more and more moved towards the railway hubs. Nowadays, the areas surrounding railway stations such as Rotterdam, Arnhem and the Hague are being transformed into multifunctional areas, combining commerce, business and leisure right around the railway platforms. Due to this way of planning, building new railway lines or doubling tracks becomes more costly and troublesome than in countries where this way of urban planning is less prevalent. The popularity of the railway network in the Netherlands thus both poses advantages as well as disadvantages.

Despite of severe limitations, the Dutch railway network is able to provide a reliable and convenient mode of transport. Punctuality is among the highest in Europe, in 2015, 94.3% of all trains arrived within the boundary of 5 minutes off schedule (NS, 2016). However, recent disruptions during the winters of 2009, 2010 and 2011 as well as other disruptions have shown how the networks' stability is prone to small delays spreading throughout the network. As the current capacity is more and more fully utilized, the robustness of the railway network suffers.

2.2 Institutional setting

For the organization of railway transport in the Netherlands, three responsible parties can be distinguished: the government, the infrastructure manager and the train operating companies. ProRail is the railway infrastructure management company of the Netherlands and therefore responsible for the maintenance and operation of the technical infrastructure of the Dutch railway network. Its core business is to provide safe access to railway tracks to the train operating companies. It does so by decree of the ministry of infrastructure and environment, which on a ten-year basis gives ProRail the license to manage the infrastructure. The Dutch Railways (NS) has a ten-year license to operate trains on large parts of the railway network. The so-called 'hoofdrailnet' constitutes the profitable core of the network on which most intercity services run. For regional decentralized routes, the responsibility lies with the provincial governments. These governments auction off the rights to operate trains to private train operating companies such as Veolia, Arriva and Syntus. These routes are often located in the outer rims of the Dutch network, for example in the northern provinces of Groningen and Friesland, the eastern provinces of Overijssel and Gelderland and the most southern province of Limburg.

Before 1995 ProRail was part of the Dutch Railways but new European guidelines meant a split-up of infrastructure management and the commercial exploitation of passenger services. According to the typology of Perry and Rainey (1988: p.196) ProRail can be defined as a government sponsored enterprise. It is a private company owned by the state and largely funded by the state while the dominant mode of control is a polyarchy of public stakeholders such as the ministry of Infrastructure and Environment, passenger representative bodies, the transport and safety inspection agency and the Netherlands Competition Authority. Every year the organization has to write a network statement in which ProRail describes how it is planning to operate the network in the coming year and at what cost. Furthermore, the organization is responsible for assigning safe train paths to train operating companies. In order to do so, ProRail has three main tasks: capacity management, asset management and traffic control. Besides that, they are responsible for expanding and improving the railway network.

Capacity management involves determining the capacity of railway sections and granting capacity on a fair and equal basis to different train operating companies. Together with train operating companies ProRail produces a train schedule on a yearly basis. For granting train operating companies (TOCs) access to the railway tracks, ProRail receives fees for every train kilometer driven on a part of the network and every stop at a railway station. Companies like NS, Arriva, and Veolia as well as freight train operators pay ProRail for the use of the network, which accounts for around one fourth of the total turnover of the organization (ProRail, 2016). Traffic control then makes sure that train paths are secure on the day itself. Traffic controllers, just like air traffic controllers, give train engineers the permission to run on predetermined routes of the railway network. In standard situations, computerized slot allocators called 'ARI' assign safe train paths to train engineers. Traffic controllers monitor these systems, which step in in case of disruptions or emergencies the ARI-system cannot handle. These systems and traffic controllers are situated in one of the thirteen regional traffic control posts. For national coordination, a central post can be found in Utrecht. Here also train dispatchers and personnel planners of the Dutch Railways are present as well as representatives of contractors. Finally, the infrastructure is maintained by the asset management department of ProRail although actual work is outsourced to contractors such as BAM Rail, VolkerRail and Strukton.

2.3 Expanding the network

Since the current railway network in the Netherlands mainly stems from before the 20th century and the late adoption meant an underdeveloped network, the sector deems the current capacity insufficient to accommodate a growing transport demand. Therefore the capacity needs to be increased: the railway sector together strives for an increase of 50% by 2020. For this goal, three basic alternatives exist (see Table 2.1). We call these: infrastructural improvements, material improvements and process innovations. Infrastructural improvements can be divided into extensions and expansions. Extensions are infrastructural measures that lengthen the total network. Examples are the railway line between Zwolle and Lelystad, built in 2011 and the Betuwe Route freight line between Rotterdam and the German border. Expansions, rather than by lengthening the network, increase the capacity by broadening the network. In recent years, many parts of the Dutch network were expanded often going from two tracks to four tracks. Examples are the partial doubling of railway tracks between Eindhoven and Amsterdam and between Den Haag and Rotterdam. Material improvements encompass all the measures focused on the trains that run on the networks with the main focus on increasing speed. Finally, capacity can be increased by using process innovations. These measures leave the infrastructure as is and are focused on improving the processes involved in running trains on a daily basis. Examples are new safety systems, decreasing halting times of trains by a more speedy boarding procedure or better traffic control concepts. In the following table the differences between these three measures are depicted.

Table 2.1: different methods to increase capacity

	Infrastructural	Material	Process
Object	Infrastructure	Infra + rolling stock	Infra + rolling stock + process + people
Cost	High	High	Low
Occurrence	Rare	Rare	Often
Timing	Long	Medium	Short
Need for space	High	Low	Low

The railway sector has opted for a focus on process innovations to increase the capacity of the Dutch railway network. This can be explained using the table above. First of all, infrastructural improvements take a long time to realize. However, capacity increases are needed before 2020 and one of the direct causes of this needed increase is the growing population, which after 2040 will stop growing and even decline (CBS, 2009). To avoid overcapacity after 2040, a sole focus on mega-projects is undesirable. Another reason for a focus on process innovations is that in densely populated country as the Netherlands, little room exits for new connections or doubling of tracks. And if possible, costs of reclaiming land and compensating adjacent inhabitants are higher than in less densely populated countries. Finally, material improvements are less effective in a country such as the Netherlands where intercity distances are relatively short and on all important railway corridors every fifteen minutes a large city is passed. Process innovations, due to their lower costs, shorter realization period and a need for less space are therefore thought to be more feasible and effective than infrastructure or material focused improvement projects.

In 2008, all relevant parties within the railway sector (ProRail, train operating companies (TOCs) and freight operating companies) developed a plan to accommodate this increase in demand. This plan, initially called 'Ruimte op de Rails' (English: space on the railway tracks) and later renamed as 'Programma Hoogfrequent Spoor' (PHS) (English: High-frequent Railway Program) promises to have delivered two improvements in the railway network by 2020: a metro-like train schedule on the most important corridors and a fixed corridor for freight with a better use of the newly built 'Betuwelijn', a freight-dedicated railway line between the port of Rotterdam and the German border at Zevenaar. OV SAAL, a project about increasing capacity on the link Schiphol Airport, Amsterdam, Almere and Lelystad is also part of this program, although decision-making is done independent from the other projects within PHS. On four corridors in the greater Randstad area of the Netherlands (the central conurbation of the biggest cities) intercity services will run in 10 minutes intervals added with regional services in 10 to 15 minutes intervals. By this, every five minutes a connection exists between the bigger railway stations on this corridor, making schedule-less travelling possible: since on average a train will depart within 2.5 minutes, passengers need not consult a train schedule beforehand. These parts of the Dutch railway network will then resemble subway networks where intervals are so short, a train will be bound to leave soon enough. The corridors that are currently planned to provide metro-like train services are: Den Haag-Eindhoven, Alkmaar-Eindhoven, Schiphol-Nijmegen and Schiphol-Lelystad (part of OV-SAAL).

However, recent disruptions during periods of snowfall have shown how the network is prone to delays spreading throughout the network. In some cases, the network partially collapsed as on many of the important links in the country railway traffic was halted. Furthermore, even during normal circumstances, the railway sector sees how the complexity of the railway system in both a social and technical sense has negative effects on the robustness of the system due to knock-on delays and technical failures in railway switches and signaling. Furthermore, the safety systems in place are evermore becoming outdated and a sense of urgency exists, both within the sector as within outside groups, to replace these systems with more modern and capacity-increasing safety systems. These observations has led the organization responsible for maintaining, controlling and expanding the railway network, ProRail, to start several projects supporting the PHS project that specifically focus on transforming the current system into a more robust one.

2.3.1 Future innovations

For the period until 2020, several innovations are planned to be carried out by the railway sector. They have in common that they focus on both technology and organization although the relative share in importance differs per case. For this period until 2020 several programs are started that embody a broad range of separate process innovations. Table 2.2 presents the most important programs:

Table 2.2: current process innovations

Project	Goal	Deadline	Linked with?
Programma Hoogfrequent Spoor (PHS)	Increase frequency to 14 trains per hour per direction on most important corridors	2015-2020	All
Robuust Spoor (RS)	Increase robustness of railway network	N.a.	PHS, KV
Kort Volgen (KV)	Decrease follow-up time between two trains	N.a.	PHS, RS
Mistral/BB21 (MS)	Install a more reliable and flexible safety system (Europe-wide)	N.a.	PHS
Spoorzone Delft (SZD)	Commission a railway tunnel without a testing period	2015	PHS

2.3.2 Programma Hoogfrequent Spoor

This program is already discussed above. Besides an ambition, the program itself entails a set of process innovations of for which most parts the specifics are now under study. For instance, a higher frequency means that less buffer exists for stopping times. Therefore passengers need to board the train more quickly and measures for this need to be taken. Also, traffic control concepts need to be tested to see if the existing ones are able to cope with higher volumes of traffic.

2.3.3 Robuust Spoor

In Japan, a country with the same infrastructure and the same technology, far more capacity is realized and with a higher punctuality (Hatch, 2000). The country thus serves as a benchmark for the Dutch railway sector (Van der Velde, 2000). One of the key differences found between the Dutch and Japanese railway operations is that the Dutch network allows for far more flexibility by having more railway switches and giving traffic controllers more freedom to alter train routes as they see fit. However, railway switches are prone to failure and flexible traffic control concepts might lead to less optimal results as traffic controllers are bounded rational agents overlooking an overly complex railway network. Therefore, in light of the PHS-program, ProRail wishes to make the Dutch railway network more robust, meaning that the network is capable to handle smaller disruptions more effectively without these disruptions spreading throughout the network as currently often happens. Process innovations that will support this are decoupling of railway corridors, new traffic control concepts, removal of railway switches and other not yet defined measures.

2.3.4 Kort Volgen

One of the important parameters of railway capacity is the length of a railway section block, the distance between two safety signals (Abril et al., 2008). In a fixed block signaling system – a system where safety signals are shown through fixed signal posts along the railway track - safety is assured by allowing only one train per section block and therefore the shorter this block is, the more trains can run on one railway corridor. For the program of 'kort volgen' or in English: short track, the idea is to decrease the length of the section blocks to a similar length as in Japan to allow more trains on the same infrastructure. Process innovations needed entail for instance more efficient placing of signaling without them becoming invisible to the train driver, information systems that allow the train driver to oversee delays of surrounding trains and other measures not yet defined.

2.3.5 Mistral/BB21

The current power supply systems as well as the safety systems are not up-to-date anymore. A higher frequency of trains on the network means that the voltage needs to be increased to 25 kV and the safety systems will be renewed. Besides a higher capacity, the program also offers opportunities for interoperability of international train services as voltages and safety systems then become the same Europe-wide. Process innovations needed are better communication systems along the railway track and on board of rolling stock, training of personnel and alteration of electrical engines on trains. Since the beginning of the program initial ambitions have been lowered as in later years ProRail started researching the financial feasibility of increasing the voltage to 3kV instead of 25kV, a voltage also applied in Belgium.

2.3.6 Spoorzone Delft

Although the commissioning of a railway tunnel is more an infrastructural improvement than a process improvement, doing this without a testing period can be seen as a new process. Currently, the connection Rotterdam-The Hague runs through the historical city of Delft and splits the city up in two halves. On top of that, no room exists for a much needed doubling of railway tracks in light of PHS. Therefore a railway tunnel is being built that follows the trajectory of the old railway line and has room for four railway tracks. Since this railway connection is one of the busiest of the Netherlands and room and money lacks for a temporary railway switch, the commissioning of the tunnel has to be done in 24 hours after the tunnel is completely built. Before that, partial tests can be carried out in a period of six months after the physical structure is built. However, in one weekend in 2015 the tunnel is connected to the existing network and within 24 hours the tunnels has to be used for the first time. Process innovations needed are training of train drivers in simulators, off-site system integration tests and better cooperative arrangements between involved stakeholders such as the municipality, emergency services and the railway sector.

2.4 From infrastructural expansions to process improvements

We expect a trend towards more process-focussed innovations within the railway sector. First of all, demand for railway transport is expected to increase in the coming years and in other countries the same volume increase goals are stated, for instance in the UK (Lovell et al., 2011). As a safer mode of transport, more energy efficient and a higher capacity (Eurostat, 2009) rail transport is deemed to be a better solution for current transport congestion problems. Furthermore the fact that rail transport causes less polution (Profillides, 2006) especially for surrounding urban areas compared to urban highways, makes it more suitable for intercity passenger transport.

Therefore, the Dutch government opts for a multi-modal approach where both marine, rail and road transport complement each other to provide the country with a flexible and reliable transport infrastructure (Ministerie van Verkeer en Waterstaat, 2008). The government report 'MobiliteitsAanpak' from 2008 elaborates on the goals for the Dutch transport network to be achieved in 2028 in light of the countries' wish to specialize in logistical services and the increasing population and urbanization. Better wet business parks (reachable by container ships), doubling of highway lanes in the wide 'Randstad' area, a higher capacity of the national railway network and efficient urban rail networks in the larger metropolitan areas of the country are the main focal points of this program. In other areas such as Nordrhine-Westphalia in Germany the Rhein-Ruhr-Express is expected to offer similar metro-like services around Dortmund, Dusseldorf and Cologne.

For achieving this, availability of financial and spatial means for building additional railway lines and doubling tracks are lacking. Especially after extending the HanzeLijn (€1bn), the Betuwelijn (€5bn) and the HSL high speed line (€7bn), large budgets are lacking for yet another round of infrastructural investments. On top of that, urban areas have coevolved with railway lines, growing just there where railway lines were able to offer fast transportation into city centres. In cities such as Utrecht, Helmond and Almere, its growth pattern can almost be perfectly explained by the railway lines that cross the cities. This development has left little room for further expansions since for doubling of tracks, many surrounding inhabitants have to compensated. A focus on process innovations, meaning leaving the railway network as is and trying to better utilize it, is therefore deemed more suitable.

2.5 Disjointed versus concerted

Since the EU regulated split up between infrastructure management and commercial exploitation, the infrastructure value chain is no longer fully vertically integrated (Weijnen & Boumans, 2007). Under EU regulation and with the opening up of transport markets for competitors, this value chain is now more unbundled. Comparing the infrastructure markets before and after liberalization, we adopt the model of Ten Heuvelhof et al (2004) cited in Weijnen and Bouwmans (2007):

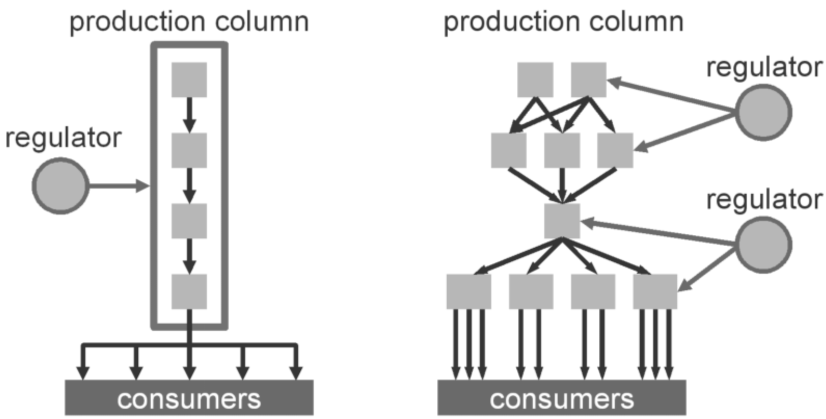


Fig 2.1: value chain of vertically integrated markets (left) and networked markets (right) (Ten Heuvelhof et al., 2004 cited in Weijnen and Bouwmans, 2007)

Figure 2.1 opens up the question to what extent the innovative efforts of all the parties involved (Dutch Railways, freight train operators, private TOCs, ProRail and governmental bodies) need to be, and can be, coordinated. To what extent may actors within the network innovate independently from other actors? We coin the terms disjointed and concerted innovation and pose that this categorization also applies for innovation within one firm. Are innovation efforts coordinated through hierarchical structures (concerted) or are efforts uncoordinated (disjointed).

2.6 Historical initiatives for systemic change

To provide an example of the railway domain that is the focal domain of this thesis, we use the Rail21 program of the Dutch railways, which was initiated at the end of the 1980s. Before that period, the Dutch railway system slowly got out of track with the broader landscape it served. Rising suburbanization, increase use of car transport and environmental concerns served as the impetus for a grand improvement program of the railway system, initiated by the then monopolist and state-owned Dutch Railways. The program, called Rail 21 – pointing to its ambition to prepare the system for the 21st century, entailed a plethora a small scale process improvements, local doubling of tracks and a few large scale network extensions. Original designers of the plan deemed it integral in that it tackled the problem of an outdated network through many qualitatively different, but coherent, measures: a new time table philosophy, new ideas about separating freight and passenger transport and usual civil engineering projects such as doubling of tracks, together with measures focused on reducing the use of cars. In itself, all separate changes could barely be classified as systemic, but put together they would systemically change the setup of the railway system. With a drastically changed setup, the system would then be better able to perform its functions such as the timely and punctual transport of people and goods. Hence, in 1989 the plan was met with much enthusiasm.

These more systemic changes, involving many different incremental changes are however highly difficult to bring about. In our example, many were in favor of implementing the whole program and one would expect a successful implementation. A few years later however, retrospective accounts tell a somewhat different story. Internal and external dynamics have caused a high skewedness of the diffusion of the innovation program into the already existing railway system. The envisioned timetable needed evermore infrastructure expansions and due to local adaptations the cost of these infrastructure expansions rose significantly. In addition, the European Union instigated a privatization of the railway sector under the 91/440 directives. This led to the separation of the Dutch Railways into two organizations: a state-owned infrastructure manager and a semi-private train operating company. By the time the program was closed in 1997, many of the intended 'soft' measures had not been implemented. Instead, the sector refocused its attention on a few large-scale civil engineering projects such as a new high-speed line, a new freight line from Rotterdam to Germany. What was left, and hence explains the skewedness, is a plethora of infrastructure measures of which some were only finished well unto the 21st century. Other more soft measures have yet to make an entrance into the current railway system. Mocking the program, Rail 21 is also remembered as Concrete 21 due to its apparent focus on purely civil engineering solutions for sociotechnical problems.

The story, albeit just one story, shows us a few important things: first of all, transforming a large complex system is even in situations where the 'multi-actorness' is relatively low – there was only one monopolist at the time - a complex project in itself. Furthermore, we see that in these cases the innovation in itself is a system, rather than an unchangeable atom, and that different dynamics occur when people try to implement these innovation system elements. Due to both dynamics within the innovation process as well as dynamics in the environment, their initially planned coherent implementation resulted in some of the elements getting implemented and some of them dying an 'early death'.

The essence of our previous example is that systemic innovation processes are not merely stories of technological change. Although technology forms an important component, the story should as well do justice to the social and institutional complexities involved. Sociable human actors need to bring about technological change and innovation is hence just as much a human story as well as a technical story. Also, the example shows that innovation processes are temporal phenomena per se and their properties can best be understood if one looks at it dynamically rather than taking static snapshots.

2.7 Conclusion

This chapter provides a first overview of the context (historical and current) in which the railway sector needs to instigate systemic change. Given examples from this context, we can typify this change as involving not only a technically complex artifact, but additionally involving complex social networks and institutional structures. These networks and structures are likely to matter in describing and analyzing the progress of an innovation process and the dynamics to the central technical artifact (the innovation itself), as the Rail21 example showed in this chapter.

3 Gaming Simulation and Innovation: a first glance

This study builds upon the premise that the design and functioning of gaming simulation is highly dependent on its context-of-use and that the context-of-use, a systemic innovation process in the railway sector, is still poorly understood. In this chapter leads for better understanding this context-of-use are uncovered, by looking at how innovation stakeholders give substance to managing incremental innovation processes and by exploring how conclusions would hold once such processes become more systemic. In addition, from what theoretically can be said about systemic innovation processes and the relations this has with potential qualities of gaming simulation, ranging from analytical-scientific to design-scientific qualities, this chapter will provide impetus for the design science approach taken in subsequent chapters. This chapter combines work on innovation processes (Van den Hoogen and Meijer, 2012) and on the relation between these processes and gaming (Van den Hoogen and Meijer, 2015).

According to Klabbers (2006; 2009) the specifics of the Design-in-the-Large (DIL), or in this case the actual innovation process, highly impact the design and usability of a gaming simulation, or Design-in-the-Small, and vice versa. As mentioned in the introductory chapter, tensions may exist between perceiving games from solely an analytical-scientific perspective and a design-scientific perspective (see for instance Mayer, 2009, Klabbers, 2009, Meijer, 2015 and Ragothama and Meijer, 2018). Are games exclusively instruments to test hypotheses about linear relations between variables, or can their instrumentality be defined in other terms? This question also strongly relates to the assumption, in this study, that processes of changing complex systems are not merely cases of a one-off decision based on an acception or rejection of a singular hypothesis on the effectiveness of an innovation. Rather, such processes are more likely to resemble ongoing negotations between stakeholders over time charachterized by a chain of smaller design decisions. In these environments, gaming simulation may lend its perceived legitimacy from stakeholders from validity of the model and game run, its usability may be defined in other terms (Reitsma et al., 1996; Barreteau et al., 2001; Ragothama and Meijer, 2018; Klabbers, 2018).

This chapter explores this assumption about the relation between innovation processes and the use of gaming simulation and the related importance (and conceptions) of performance indicators such as validity and usability. This part of the thesis investigates two exemplary perspectives to the relation between DIS and DIL based on a first empirical glance on the use of gaming simulation in the Dutch railway sector in the period between 2010 and 2012. We look at how innovation stakeholders in practice give meaning to the relation between DIS and DIL. We do so because the remainder of this thesis, including the theoretical framework and methodological considerations in chapters 4 and 5, builds strongly upon the conclusions made from this first empirical investigation.

The first perspective on the relation between innovation and gaming stems from the analytical sciences, positioning games as pure experiments intended to deliver valid and reliable statements about the (linear) relation between variables. It is for this quality that such conceptions of games as experiments fit more linear models on innovation, viewing such processes as moving from experiment to experiment until final implementation (see for instance Thomke, 2003). The second perspective has a more social constructivist stance, positioning games as means to collectively envision (and design) new avenues for improvement within a broader context. This context influences policy making processes (parallel to the conducting of a gaming simulation session). The latter perspective then resembles a design science approach, where credibility and usability are parameters by which one can assess a gaming simulation, rather than validity and reliability (Klabbers, 2009).

This empirical chapter precedes a more elaborate theoretical framework brought forward in the next chapter. We do so since the value of gaming simulation to innovation processes was only uncovered over the course of the research project, spanning a period of 6 years. In this period the learning process of us as scholars and designers of gaming simulation involved acknowledging that the conception of the DIL, i.e. how within the sector innovation was conceived and organized, and the conception of the DIS, i.e. how the client (and we initially) deemed a gaming simulation had to be designed, were indeed highly interrelated but initially started from an apparent false premise. This premise was that innovating in the railway sector is mainly a techno-rational endeavour where the increasing of knowledge and decreasing of uncertainty over time is key. Subsequent chapters will provide, based on literature and empirical research, a framework to differently, and in more detail, conceive innovation processes (DIL) and the way games should be designed accordingly (DIL).

For now, this chapter empirically depicts how a railway sector conceives innovation and how such processes are organized within an organizational context. This chapter will show that this conception is characterized by its linear and techno-rational nature. It proposes that such conception by innovation stakeholders from within the sector will become incompatable with the nature of systemic change: its inherent dilemmatic and paradoxical nature given the social and institutional complexities that add to the mix. This conception of innovation processses however also explains the conception of the role of gaming simulation initially perceived solely from the analytical sciences: games are simple machines delivering valid causal claims about the relation between variables. This chapter will show that gaming simulations as such pure experiments are inherently problematic and positioned in their context-of-use will likely be suboptimal: validity and usability measures are in some instances, as the analysis in this chapter shows, mutually exclusive. However, we note here that involved stakeholders usually tend to lend legitimacy to the use of models and gaming simulation via the perceived validity of the outcomes (Reitsma, 1996; Meijer, 2015).

This chapter brings forward the aforementioned claims about gaming simulation in innovation processes in the railway sector by answering four reseach questions:

- How do stakeholders involved in introducing gaming simulation conceptualize innovation processes and how do these processes manifest in real life?
- Given two models on the relation between DIS and DIL, what theoretically induced qualities should a gaming simulation have?
- Based on a range of gaming simulation sessions, how does gaming simulation perform on these quality measures?
- What model of the relation between DIS and DIL is likely applicable to systemic innovation processes?

These four questions are based on two observations: 1. Gaming simulation literature assumes that conceptions of context-of-use and game design are highly interlinked and 2. The client of the initial games designed for the railway sector was the innovation department of ProRail. Hence design, employment and use of games in the railway sector involved the same department also involved in bringing about innovations. The following questions are then relevant since this locality of the use games may provide leads for a more fruitful conception of gaming simulation if a) the context-of-use changes as proposed innovations become more systemic and involve a more diverse range of stakeholders and b) the use and value of gaming simulation and its design changes accordingly. The chapter will first adress two models on games' embedding in innovation processes. The empirical part will then be a case study on incremental innovations in the railway sector and an analysis on how games function when looked from either of the two models.

3.1 Two theoretical models on the DIS-DIL relationship

Although there is an extensive range of literature on both innovation and gaming separately, little scholarly work exists on the juxtaposition of the two. For our initial analysis of the empirical findings we therefore use two models, one from the product development literature and one from the game sciences and translate this to the specific domain of this thesis. We use these models to see which one best applies to the current situation in the railway sector, normatively and descriptively and which one is best suited to understand and improve systemic innovation processes and the use of gaming.

The first model (see figure 3.1) stems from the innovation and product development literature that focuses on the intelligent use of experimentation in innovation processes. The model is based on a cognitive model of innovation that involves an entity, e.g. the designer, constantly generating and testing design alternatives (after Simon, 1969). It is cyclical in nature in that innovating involves a trail-and-error process where designers make an educated guess about where a solution may lie and test this using experiments and prototypes (Thomke , 1998; 2001). Empirical research into this model shows how in the case of product development such problem-solving indeed takes this form (Clark and Fujimoto, 1991; Smith and Eppinger, 1997). The core of this model is thus that within the innovation process itself an iterative process takes places where innovation stakeholders conceive of an innovation, build and run an experiment to test this innovation and subsequently have the analysis of the result impacting the further conception and design of the innovation.

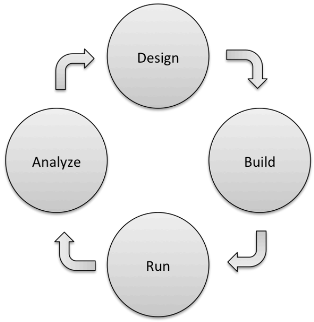


Figure 3.1: a cyclical model of problem-solving within design processes (Thomke, 2001)

The second model conceives gaming in its context-of-use, in this case policy development, differently. Rather than one single process in which gaming is embedded, the design and use of gaming simulation occurs parallel to the process in which it is more indirectly embedded than in the cognitive model of Thomke (2001). This model by Klabbers (2009) differentiates between a macro cycle (how the design in the small relates to the design in the large process and the micro cycle (the design of the game itself). It is worth noting that this model includes the notion that other factors beyond the dynamics in the gaming experiment itself impact the policy making process and that the process itself has a strong temporal feature (depicted by the arrow in Figure 3.2). The notion of micro and macro cycles and the separation of a game design process and a policy making process then make for a more complex picture of the use of gaming simulation when translated to its actual use in innovation processes.

Comparing the two models they seemingly originate from two different scientific traditions. Thomke's model sees innovating purely as a cognitive problem-solving process where 'problem-owner' and 'game designer' belong to the same entity and their respective activities, i.e. designing and testing, occur within the same process. Although iterative and cyclical, the model therefore is also highly linear: designing innovations and designing games to test them are activities always following each other in a specific order. In addition, the model perceives innovation as an isolated and atemporal process: in understanding such processes the environment and the timing of multiple processes does not matter. Within this model the approach taken to the building, running and analysis of the experiment is similarly stemming from an analytical science approach: the design of an innovation is seen as a independent variable and its linear relation to a dependent variable (or performance measure) is tested. Validity and reliability measures then describe the quality of the experiment.

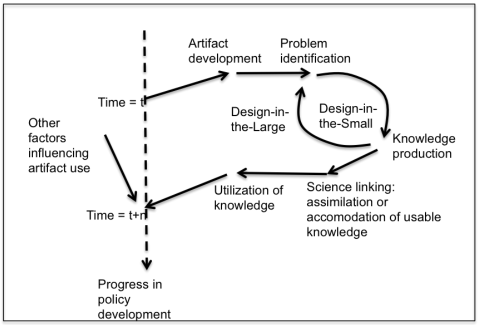


Figure 3.2: game design (micro-cycle) and its relation to a context-of-use (macro-cycle) (Klabbers, 2009: 291)

Klabbers' model then adds to this picture the inherent social and constructivist nature of innovation, i.e. people desiging future solutions, and how designing innovations and designing games within such processes can be separated over different actors and over different subprocesses. This makes way for dynamics not solely being caused by cognitive processes, as is the case for Thomke's model, but also by processes of social, political and institutional factors. In addition, such pressures may be endogenous to the model, for instance when defining a problem or deciding on a method, but also exogenous to the model. It then is worth noting the temporal nature of processes which become relevant when outside of (and during) the parallel micro-cycle of game design other factors affect the policy making process as well and these highly interact. In this conception of the relationship between gaming simulation and an innovation process, the game is not merely an experimental machine, as valid and causal claims are impossible: the reference system a game is trying to model is a moving target and true validation is therefore impossible. Table 3.1 provides an overview of the differences between the perspectives.

Table 3.1: two perspective on the use of gaming simulation in innovation processes

	Analytical Science	Design Science
Function	Test hypotheses	Improve change processes
Assessment on	Validity and Reliability	Usability and Credibility
DIL vs DIS	DIS linearly follows DIL and vice versa -> they are part of the same cyclical model	DIS and DIL are separated (and temporal) cycles that interact in more complex ways
Underlying model of innovation	Linear model of innovation, cognitive problem-solving, Quality of innovation can be determined objectively, from the outset No relevant context, innovation is an isolated process Timing is irrelevant	Non-linear model of innovation, Socio-institutional dynamics, Parallel processes Quality of innovation is socially constructed, during the process Context is relevant Timing is relevant because of multiple parallel and interdependent processes

3.2 The use of these models in the Dutch railway sector (2009-2012)

The way this chapter intends to explore the use of these models on innovation and the use of gaming simulation is by looking at the way innovation processes currently occur within the sector and the way different gaming simulation sessions would subsequently be designed. The following three steps describe the structure of this chapter:

1. Analyzing current innovation processes in the railway sector
2. Analyzing gaming simulation based on the two models and comparing games with computer simulation and live-tests.
3. Concluding on which model will better help in theorizing, understanding and improving systemic innovation processes and the design of games.

3.3 Analyzing the Design-in-the-Large (DIL) processes in the railway sector

Although systemic innovations are rare, their individual components (incremental changes) are not. The organizational department involved in introducing gaming simulation was also involved in several of these incremental innovations. Given that we wish to see how these stakeholders conceptualize and manage innovation processes, we first studied a range of incremental innovation processes inductively. Knowing that these are not systemic and knowing in basic terms how incremental and systemic innovations differ, we could also deduce what interesting phenomena, such as functions and dysfunctions of gaming, to expect when these incremental changes do become systemic. This impacts both what theoretical frameworks would do justice to the qualities of such innovation processes as well as how ideally gaming simulation processes (its design, employment and use of outcomes) need to be organized. As our main goal is to give a description of current practices at ProRail, we choose the case study method as most suitable for answering our questions (Yin, 2013). We conduct an in-depth study on the phenomena we are interested in.

We looked for respondents mainly in the innovation department of ProRail, although some of the respondents also worked for other departments involved in innovating as well. The selection criterion was that they had been involved in small-scale improvement projects in the last two years and as such could tell in detail how they executed these projects. The assumption here is that in their reasoning about their innovation projects and the eventual use of prototypes, tests and simulations, a broader and shared understanding of innovation processes would come to light, that speaks for the culture of the innovation department.

The commissioning of an old railway line between Nijmegen and Kleve is included in the analysis to allow for a comparison with a more infrastructure-focused project. We conducted eight open interviews during which the respondent was asked to describe several projects undertaken. Interviews took about one hour and were taped using a recorder. We did not give a detailed description of the interview and its goals beforehand. The interview only started with the question what projects had been undertaken in the last two years and if the respondent could describe them. This ensured an open and free interview in which respondents addressed topics as they saw fit.

We made additional field observations between March and August 2011 during railway simulation sessions, project meetings, as well as site visits. We have visited two regional traffic control centers in The Hague and Rotterdam as well as the national traffic control center (OCCR) in Utrecht. A triangulation of data ensured the validity of the final model about decision-making on incremental innovation. Only interviews could lead to a bias as respondents might rationalize their decisions post-hoc (Child et al., 2009). We were able to look if propositions from the interviews could be corroborated by the field observations we made ourselves.

We followed to some extent a grounded theory approach in which we first gathered data and based on this data build a theory, using the method as proposed by Strauss & Corbin (1990) in which theorizing is only done after data collection. We were completely new to the domain of railways and therefore could fulfill the unbiased observation requirement. The interviews and observations provided a range of projects to study. Seven projects were discussed to such an extent they allowed for analysis. Of these seven, six involved a process innovation and one involved an infrastructure project. By doing so, we are able to compare the decision making process on small-scale process innovations with the decision making process on larger-scale infrastructure extensions. Tables 3.2 and 3.3 present the overview of the interview data in condensed format.

In the data we looked for patterns that explain the decision-making and success of process innovations in railways at the project level. For every pattern found, the full transcripts of the interviews were recalled and checked for correct interpretation of the pattern. The following sub-sections present three propositions formulated on 3 patterns found that remained when checked in the transcripts.

Table 3.2: different innovation projects compared (1/2)

	1. NAU	2. ETMET	3. GPS tracker
Problem	Disruptions at Utrecht, central node of the network, spread throughout the network	Feasibility of frequency increases on main corridors unknown	Exact location of trains unknown on certain parts of the network, problematic travel information
Goal	Make Utrecht more robust to delays	Determine feasibility of 14 trains per hour per direction on Eindhoven-Amsterdam corridor	Provide traffic controllers with accurate and actual position of trains
Contested goals?	Partly, for NS it was also about lowering work pressure of train controllers	n.a.	No
Client	Prorail Board	ETMET task force	No direct client
Stakeholders	Many	Many	Few
Restrictions	No changes to IT- and safety systems	Test on only one of the four corridors	No changes to IT- and safety systems
Inhibiting factors	Such measures always involve multiple parties, parallel developments and different lead times	Limited control over behavior of operational personnel	Limited control over behavior of operational personnel, GPS should be fail-safe
Solution	Decrease amount of trains crossing each others path	Increase in frequency is feasible but is very unstable	Place GPS on board
Result	Implementation of corridor concept, restricting the use of rail switches	Proof-of-concept	Proof-of-concept
Remaining uncertainties	When to not use the corridor concept? What is the role of national traffic control?	Feasibility once management attention fades away	Use of the GPS by operational personnel; Implementation time of proof-of-concept
General remarks about organization	Incident-drive culture, short term focus, compartmentalization	n.a.	Hierarchical, top-down decision making. Focus on own departments

Table 3.3 different innovation projects compared (2/2)

	4. Rail-wheel lubricant	5. Countdown	6 Spreading Passengers	7. Nijmegen- Kleve rail link
Problem	Noise pollution due to rail-wheel friction	High variance in dwelling times on railway stations	High dwelling times due to uneven spread of waiting passengers on station platforms	Increase in traffic demand between Nijmegen and Kleve (Germany)
Goal	Decrease friction and noise pollution, later goal added: reduce wear-and-tear	Decrease variance in dwelling times	Decrease dwelling times by speeding up boarding of passengers	Re-install old railway line
Contested goals?	Partly. NS is not responsible for noise, but is responsible for wear-and-tear and passenger comfort. These goals were added	No	No	Yes. Ranging from building a tram to linking Weeze airport to Dutch network. Municipalities feared freight traffic
Client	ProRail Asset Management	n.a.	ProRail Traffic Control	EUregio Rhine Waal
Stakeholders	Few	Few	Few	Many
Restrictions	No interference with daily operations, no impact on braking distance	No changes to safety systems	No permanent changes to station platforms	Subsidy of EU directed to using light rail
Inhibiting factors	Misaligned incentives. Cooperation of NS was necessary but gains were mostly for Prorail	Limited control by Prorail over behavior of operational personnel of NS	No interference with daily operations,	Misaligned goals, different parties wanted different transport modes
Solution	Use of already existing onboard lubricant to also use for rail tracks	Connect a device to safety system that tells train drivers when to start departure procedure	Provide passengers with dynamic information on length of train	No solution
Result	Proof-of-concept	Proof-of-concept	Proof-of-concept	No result
Remaining uncertainties	Feasibility for lines with multiple trains from different companies	Train driver behavior	Ability of NS to timely provide accurate train length	N.a.
General remarks about organization	Risk-averse culture, compartmentalizatio n	n.a.	n.a.	Focus on procedures and solutions, not on the underlying problem

3.3.1 Proposition 1: projects are mainly focused on delivering a proof-of-concept

As shown in Table 3.2, nearly all projects ended up with proof-of-concept. Only NAU project ended up in an implementation. For instance, spreading passengers along the platform was done at a small railway station to see if the measure worked at all. One respondent mentioned that usually new measures are tested in rural parts of the country and if results are positive, tests are done in busier parts of the railway network. We have seen concrete evidence of this at the project on reducing railway friction and the noise pollution this causes. The chosen alternative was one where a lubricant machine would be placed under the trains that directly lubricated the railway tracks as the train traversed the network. This measure was first tested on a regional line with a smaller train operating company before ProRail went to the larger Dutch Railways with the solution. Through the use of a proof-of-concept, either tested in a simulated environment (projects 1 and 2) or in real-life (projects 1 to 6) project managers determined the effects a possible solution has and to what extent it helps in solving the central problem. The first reason for a focus on proof-of-concepts is that project managers needed to be convinced that their solution would be feasible and effective. Without a proof-of-concept, project managers were unsure of how in particular the human elements of the railway system (passengers, operational personnel) would react to the measures. In project 7, the commissioning of a railway line, these substantive considerations were far less prominent. Here, the decision making process was far more about multiple parties achieving their own objectives. Besides that, mere practical arguments can be given for not focusing on a proof-of-concept in real-life for a complete railway line.
Besides proof-of-concepts as a means to convince oneself of the effectiveness of a measure, they also served to convince other parties not yet involved in the decision making process. A partial goal to test the onboard lubricant on the lines operated by Connexxion was to convince NS that this measure was effective in reducing noise pollution. ETMET was a project specifically targeted to achieve a proof-of-concept. The main question during the project was: are we able to increase the frequency on the corridor Amsterdam-Eindhoven given the current infrastructure?

In many cases the proof-of-concepts showed the feasibility of the alternative. The onboard placement of a lubricating machine on the decentralized line between Amersfoort and Ede-Wageningen proved to be effective in reducing noise pollution and informing waiting passengers about the length of a train proved to be effective in spreading them across the platform to ensure a quicker boarding procedure. In other cases, the proof-of-concept showed that the solution would still be infeasible. For instance, the live-test of increasing the frequency to 12 passenger trains per hour on the corridor Amsterdam-Eindhoven showed that the current system is not able to cope with moderate disruptions on such densely utilized parts of the network.

3.3.2 Proposition 2: concessions are made to arrive at a proof-of-concept quickly

Since project managers cannot predict the effectiveness of a measure, a proof-of- concept proved to be valuable. To quickly come to such a proof-of-concept, we saw how project managers early on in the innovation process placed large restriction on the alternatives they would consider. If we define the range of considered alternatives as the 'design space', project managers reduced technical complexity and social complexity by minimizing this design space. For instance, to decrease the technical complexity, solutions that involved interfering in safety systems and IT-systems were purposefully neglected in projects 1, 3 and 5. Respondents explicitly mentioned that better solutions were most probably forgone but that limiting the design space leads to faster results in a proof-of-concept. Project 1 purposefully followed the measures NS already took to make the Utrecht more robust. In case of project 2, it was chosen to only focus on one corridor (A2) rather than the total four or five that are planned to have this frequency increase. In case of project 6, the solution was designed to work on a railway station with only two platforms, where only one specific regional service stops and under the premise that NS was able to tell in advance the length of the trains. In that way, additional technical complexity of last minute platform changes, different services and unpredictable train lengths was avoided.

Besides a reduction in technical complexity we have seen how innovation managers also reduced social complexity by not involving too many departments early on in the process. Many respondents mentioned that the organization can be characterized by a plethora of self-reliant islands of many different disciplines. This compartmentalization is perceived as effective for governing a railway system as long as procedures would ensure a close coordination of activities. However, for innovation projects this compartmentalization proved to be an inhibiting factor. Many respondents mentioned the need for approval of different departments as a burden. Overlapping change initiatives sometimes ran in parallel to each other. We have observed ourselves how comparable simulators were developed in different departments at the same time without the departments being aware of each other's activities. In the innovation process, respondents pointed specifically to the different lead times each department had and pointed to the extra demands a department would have if a measure would also involve them. By decreasing this social complexity, project managers were able to more ensure a timely proof-of-concept.

This reduction in social and technical complexity by minimizing the design space has led to a manageable innovation process towards a proof-of-concept. In almost all cases we have seen how the problem formulation and goals were uncontested. In projects 1 and 4 goals were contested but only after the first proof-of-concept was made in a simulated and real-life environment respectively. Furthermore, distinct phases in the innovation process could be distinguished and these phases followed each other sequentially rather than iteratively, simultaneously or otherwise. In project 7, the participation of multiple parties early on in the decision making process has led to the opposite observation. Problem formulations and goals were highly contested and changed frequently during the decision making process.

Where the first goal was to install a light rail link, one municipality reformulated the goal. According to them, the goal was not to build a light rail link but rather to provide a public transport link for which a bus connection would also suffice and would diminish rail-related noise problems for the municipality.

3.3.3 Proposition 3: concessions lead to uncertainties about the effectiveness of a proof-of concept

What all these proof-of-concepts had in common is that they showed the feasibility or infeasibility of the solution in a closely defined part of the railway system. The restrictions project managers placed on their design space abled them to develop a proof-of-concept but these restrictions inhibited to a large extent the external validity of their findings. In all cases where a proof-of-concept was made, projects managers were uncertain about the effectiveness of their measure if it would implemented in the total railway network.
For instance, showing information about the length of a train proved to be effective but only under the assumption that NS could guarantee a predictable schedule about the length of the trains. Furthermore the solution was tested on a station with only one platform per direction, leaving out the possibility for last minute changes in departure platforms. It therefore remained highly uncertain to what extent the solution would be effective on a railway station with twelve platforms, serviced by different train operating companies and with a dynamic length of trains. The lubricating machine was effective but only on a part of the network where just one type of train ran and its effectiveness on parts where multiple types of trains with different speeds is still questioned by the involved project managers.

The many uncertainties about the effectiveness of these proof-of-concepts once implemented in the complete railway system caused almost all projects to end up without a clear implementable result. Project 1 is one the exceptions to this finding as the project resulted in an implementation of the alternative that increased the robustness of the central node of the Dutch network to minor disruptions. The proof- of-concept during this project was tested in a simulated environment that through the use of gaming elements made less restrictive assumptions than the other projects for which the proof-of-concept was tested in real-life.

3.3.4 Essence of current DIL processes

Our results show that innovation stakeholders consider the railway system as technically complex and that the configuration of decision makers responsible for different parts of this system, adds to the social complexity. Furthermore we have shown that of all the innovation projects only a handful were implemented and that this leads to incremental steps being taken rather than radical or breakthrough innovations. Existing research on railway systems and network-based infrastructures in general acknowledges that infrastructures are characterized by technical and social complexity (Herder et al., 2008) leaving little opportunity for one actor to solve the problem on its own (De Bruijn and Herder, 2009). The complexity of railway systems therefore favors incremental innovations (Geyer and Davies, 2000; Bontekoning and Priemus, 2004; Geels, 2005).

Whereas the abovementioned literature on innovation in networked infrastructures puts a primacy on structural features as a causal factor for incrementalism, our first analysis adds that actors themselves strive for incrementalism as well. Using the two models on the linkage between Design-in-the-Large and Design-in-the-Large, the innovation managers appeared to conceptualize innovation according to Thomke's linear model: they actively sought to make the process behave like a linear and sequential process, as in the model, and subsequently design, ran and analyzed the proof-of-concept from an analytical science perspective.

We have seen how by limiting the technical and social complexity beforehand, innovation managers were able to decrease the amount of political behavior in decision-making. Rather than a network of actors with different goals, incentives and opinions about valuable information and the decision itself (De Bruijn & Herder, 2009) we see that a phased innovation approach can be distinguished in the process towards a proof-of-concept and that such concept can be tested using an analytical-scientific approach. By purposefully limiting the 'politicalness' inherently associated with technically and socially complex issues, innovation managers were able to quickly deliver a proof-of-concept and tested in a pseudo-classical experimental setting. The restrictive assumptions under which a proof-of-concept is valuable, given the inherent flaws of simulations and live-test as analytical-scientific tools, cause the uncertainties about the feasibility of the proof-of-concept once implemented in the complete railway system to remain high. Only one of the studied projects therefore achieved implementation.

This first analysis points to two interesting phenomena: firstly, it is apparent that innovation stakeholders not only conceptualize innovation processes using more linear models, they actually work to realize such linear models in real life. Regarding the use of experiments, broader than simply gaming simulation, we see in these cases that the experiment always follows already drawn up hypotheses (or design alternatives) and that the experiments' success or failure is described in terms of its validity.

Secondly, regarding theoretical frameworks to better understand (and improve) systemic innovation processes, these need to deal with both sides of the agency-structure debate (see Giddens, 1979) in sociology research: innovation process dynamics are both driven by structural features (networks of interdependent actors) as well as by individual human agency (purposeful behavior by individual actors). In addition it points to the inherent non-linear nature of systemic innovation processes. The studied processes were linear but project managers actively sought to make it more linear by keeping the systemicity of their innovation low. High restrictions in technical and social complexity are elements that contradict with everything a systemic innovation is. We therefore have to find frameworks that deal with the assumingly non-linear nature of systemic innovation processes and adopt a methodological approach by which the role of gaming can be investigated.

3.4 Analyzing the Design-in-the-small (DIS): gaming simulation

Whereas the previous analysis looked solely at innovation processes, the second analysis introduces the value of gaming simulation. The method of gaming simulation is especially relevant since the advent of gaming simulation runs parallel to the advent of the systems sciences and the need to better design, control and change complex sociotechnical systems. The specifics of complex sociotechnical systems then add to the problem: a complex system is inherently incomprehensible to one expert (Rycroft & Kash, 1994). Since the work of Simon (1955; 1959) introduced the concept of bounded rationality, many scholars have shown how the assumption that human beings are rational agents cannot hold, due to biases in their processing of information (see Hogarth & Reder, 1987 for an overview of biases). Experiments on handling even the simpler forms of "complex systems" show how the complexity of a system correlates negatively with the performance of managing such a system (Diehl & Sterman, 1995). As the problems of understanding and managing complex systems started to be seen as "wicked" problems, their solutions were no longer merely seen as an exercise in mathematical optimization. To understand complex systems, a more expansionist and holistic approach, rather than an analytical and reductionist approach, was needed (Ackoff, 1974).

Simulation as a third way of doing science (Axelrod, 2003), alongside inductive and deductive reasoning, proved to be a suitable method to study systems holistically, doing justice to both microscopic simplicities and macroscopic emergent complexities. Bratley, Fox and Schrage (2011: 9) define the act of simulation as "driving a model of a system with suitable inputs and observing the corresponding outputs." Hence, simulation involves both modeling, building an abstract representation of reality, and experimenting, manipulating the underlying parameters of this model. The use of simulation to study complex phenomena is said to have started with the advent of the digital computer and the first use of the ENIAC computer to model thermonuclear reactions in the wake of the Manhattan Project (Winsberg, 2010). Since then, simulation has found its way into diverse fields, such as physics, meteorology, operations research and the social sciences (Axelrod, 2003).

Gaming simulation follows a similar logic, where researchers make a simplified model of a reference system. Gaming is ideally suited for this purpose given that games in essence are structurally isomorphic to sociotechnical reference systems. Games, as do sociotechnical systems, consist of players, rules and resources corresponding to the make-up of social systems (see Table 3.3, Klabbers, 2009: 136).

Table 3.3: Gaming as isomorphic to social systems (Klabbers, 2009: 136)

Games	Rules of correspondence	Social Systems
Actors (players)	<->	Social organization, system of interactions, collective network
Rules	<->	Laws, customs, code of conducts, corpus of assertions
Resources	<->	Resources (renewable/non-renewable resources, infrastructure, goods, flow of matter, etc.)

With such model of a reference system one could offer a stimulus to the game and its players, after which one transfers the conclusions to the real-world system (Peters, Vissers & Heijne, 1998). As such, gaming simulation has been employed to study real-world systems in which human behavior forms a vital part, starting with applications in warfare and military logistics (Mayer, 2009). In this tradition, adding “gaming” to simulation was a matter of adding humans to a simulation run to increase validity, distinctive for an analytical-scientific approach to the design and use of gaming simulation. The game itself would remain a black-box, or a so-called trivial machine, where the game merely functions as an a-historical a-contextual translating device, transforming an input variable x into an output variable y (see Von Foerster, 1984 and Klabbers, 2009).

It soon however became apparent that a myriad of incongruent incentives and perceptions and the inherent chaotic properties of feedback systems such as those found in human organizations and sociotechnical systems caused the wickedness of the problems involved in such systems. This recognition led to a sense that any statistical relationships would be dubious and long-term planning impossible (Stacey, 1995). The participatory variants of simulation, ranging from simulation with gaming elements to complete free-form games, offered the chance to create consensus between decision-makers through the multilogue mode of communication. In these games, people with different perspectives engage with each other simultaneously (Duke & Geurts, 2004). Gaming simulation, compared to computer simulation, became increasingly consensus-oriented rather than scientific-oriented (Geurts & Joldersma, 2001). In this sense, gaming came to be seen as “the participatory successor of computer simulations” (Geurts & Joldersma, 2001: 305). Research has shown the positive effects of gaming simulation through providing decision-makers with a common language (Joldersma & Geurts, 1998), fostering ideas (Duin, Baalsrud Hauge, Thoben & Bierwolf, 2009) and organizational learning (Klabbers, 1993; De Caluwé, 1997).

Gaming simulation can thus be employed as either a research tool or as a design instrument for policy-making (Peters, Vissers & Heijne, 1998). Gaming simulation serves as a tool for bringing diverse insights together as well as a tool that allows alternative solutions to be envisioned in a safe environment (Hofstede, Caluwe & Peters, 2010; Joldersma & Geurts, 1998; Kriz, 2003; Meijer, 2012). In the case of transforming systems, both of the benefits of using gaming simulation might play a large role. The recent “rediscovery” of gaming simulation as a concrete decision-support tool has made the distinction in value between simulation and its gaming counterpart rather opaque.

3.4.1 General design framework for gaming simulations

Within Thomke's framework, designing a gaming simulation would need to strongly resemble the design of a classic scientific experiment. In such case, one would compare a system with and without an innovation in which a treatment group is evaluated before and after the treatment (the classical 2x2 reseach design). Within Klabbers' framework the designing of a gaming simulation would involve different rationales, depending strongly on the specifics of the context-of-use, mainly focused on credibility and usability (Klabbers, 2009). Games can thus be designed from an analytical science perspective and a design science perspective. Although gaming simulations differ in forms and purposes depending on the perspective used, still a set of fundamental design characteristics and parameters can be distinguished. Figure 3.3 depicts a meta-framework that includes gaming simulations for research, training and policy purposes, specified to analytical science and design science (slightly adapted from Meijer (2009)):

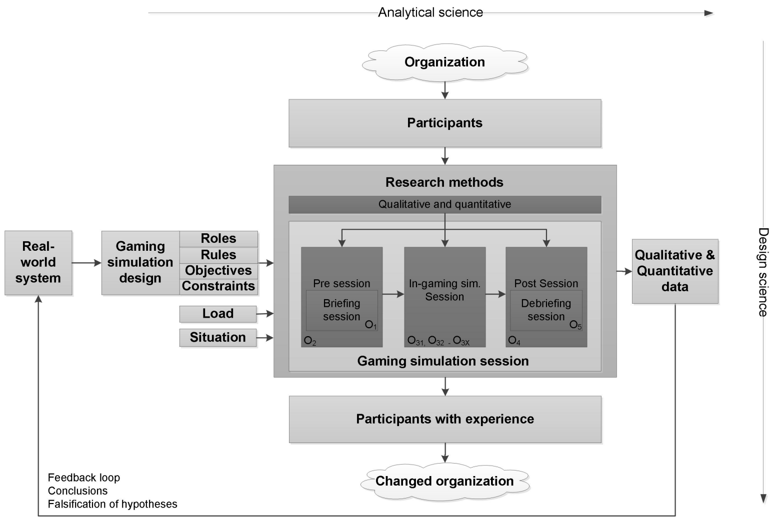


Figure 3.3: Meta-framework for design and analytical science with research and evaluation procedures included (slightly adapted from Meijer 2009).

Following the framework in Figure 3.3, components within the real world provide input for the gaming simulation design aspects. These design aspects are related to the roles, rules, objectives and constraints of the gaming simulation with parameter settings, such as load and external influencing factors. Roles within gaming simulations can exactly match the roles of participants in the real-world environment or rather abstract representations (i.e. a fantasy role). Rules refer to behavioral limitations in the reference system or artificial constructs in what is allowed or forbidden within the simulated system. The nature of objectives needs to be determined, to include individual and/or team goal(s) that are (implicitly) present in the reference system. Through the constraints, the range of actions that participants can take is limited within the gaming simulation. Additionally, the value of the variables in the design of the gaming simulation (load) and external factors that are present in a gaming simulation, set the parameters of the gaming simulation. The abstraction level with regards to physical elements, which can be operationalized by the level of similarity and accuracy and use of isomorphism with the reference system, is determined by the choice of one of the two scientific approaches.

For example, the emphasis on the validity of the gaming simulation influence these design choices, which differ between design and analytical purpose. Peters et al. (1998) identify four types of validity for gaming simulations that are used for research, policy and educational purposes. Psychological reality refers to the perceived realism of the gaming simulation environment (i.e. simulated system). Structural and process validity refer to the degree of isomorphism in the simulated system with regards to the underlying structure and resulting processes in the referent system. Lastly, predictive validity denotes the degree to which the outcomes of gaming simulation correspond to historical or future outcomes in the reference system. It is self-evident that gaming simulations that serve the purpose for research require high validity levels on all four validity types, followed by educational gaming simulations which have a lower priority on predictive validity, and policy gaming simulations whose performance relation to validity is unknown. In the latter case, given its constructivist purpose, gaming simulations can be measured according to its credibility and usability to the involved stakeholders in the context-of-use (Klabbers, 2009). In that case, validity may play a role in increasing the credibility in the instrument perceived by relevant stakeholders and via this quality improve the process in which it is used. In such case the relation between validity and quality is at best indirect and mediated by other more relevant qualities of gaming simulation for policy making.

The next block in the framework describes the gaming simulation session, with a particular focus on the qualitative and quantitative data that is acquired to feedback the participants for training and policy (design science) games or to collect data for hypothesis testing in research (analytical science) games. In case of an analytical science approach, the research design and methods need to be carefully aligned and integrated. Such gaming simulation sessions are usually consisting of a pre-session that can be separated in a briefing session, in which one or more participants are briefed on the session, and a window for measurement before the start of the session. During the session usually more qualitative and observational methods are used, followed by a possible measurement directly after the end of the session, and a final debriefing in which the participant(s) reflect about their experiences in the game session.

Gaming simulations exist in different forms, e.g. from high-tech individual human-in-the-loop simulator alike environments to low-tech multi-actor gaming simulations. The latter uses isomorphic elements, in which the information systems are made more abstract e.g. trains are represented by sponges or pegs. Train traffic operators take part in the gaming simulation in their own professional role, as long as the game is intended as a pure experimental research instrument. In such instances, all necessary information is provided to the operators to make similar decisions as in their real work environment. In other instances, for instance to conceive in-game different future solution or policy options, different demands can be placed on the operators realistic behavior and the information presented to them.

As mentioned earlier, different methods can be applied to test innovations in complex systems. Computer simulation as well as gaming simulation are both methods of simulating a reference system, each with their own properties and related strengths and weaknesses. Different purposes guide the development of both types of simulations. For computer simulations that are used for research, it is necessary to look into the process of simulation and conducting the research, which include the development of the model, the data analysis and the feedback of the results to others (Axelrod 2003). However, this is also the case for gaming simulations. In essence, gaming simulations experiments (or direct experiments) follow more or less the same research process as computer simulations (or thought experiments) (Axelrod 2003, Sterman 1987).

3.4.2 Assessing gaming simulation in the analytical sciences

Within the analytical sciences, the quality of gaming simulation is determined by the extent to which it is able to allow for conducting valid and reliable experiments, regardless of the impact of the resulting causal claims for the innovation process. Lo et al. (2013) provides a framework to assess the degree to which gaming simulation is able to provide valid causal claims in which we distinguished different dimensions of the concept of validity.

In Figure 3.4, the research process of both simulation and gaming simulation is presented, when purely perceived from the analytical sciences. The model focuses on three levels:

1. To model or create a simulated system that represents the reference system
2. To select valid simulation strategies or facilitate natural behavior by participants whilst controlling the research environments for confounding factors
3. To identify and obtain valid and accurate outcomes of the system that need to be translated to clients or researchers. This is in line with the process where a problem entity is translated into a computerized model through a conceptual model (Sargent, 2005).

In Figure 3.4, the three levels are accompanied by a set of validity challenges that have certain assurance for the following level. In order to have a valid simulated system, the *external validity* (the degree to which the findings can be generalized (Campbell and Stanley 2015)) needs to be assured. To confidentially make causal claims from the collected data (also defined as *internal validity* (Zechmeister, Zechmeister and Shaughnessy 2001)), the session needs to be controlled for internal validity threats. Finally to draw conclusions based on the used research methods, these research methods need to be assured of a high *test validity*.

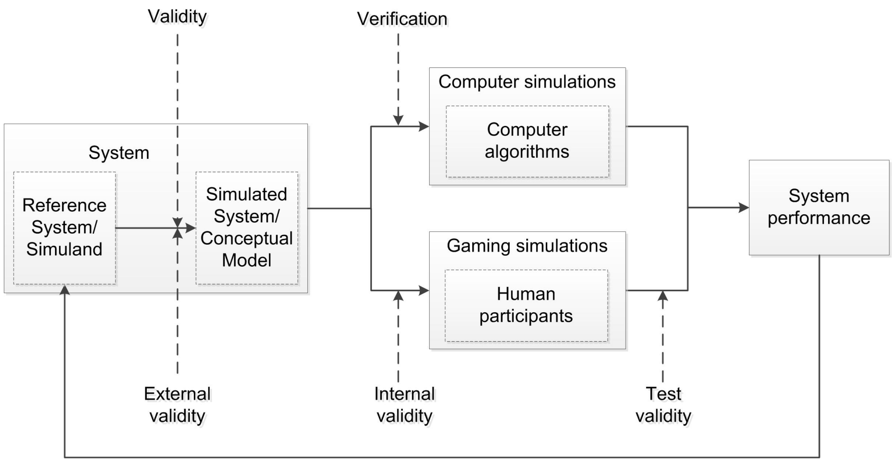


Figure 3.4: Three-leveled challenges in the research process of computer and gaming simulation environments (Lo et al., 2013).

In comparing computer simulation with gaming simulation, especially internal validity and test validity issues become significant. As a closed system, and thus lacking the problems of confounding factors, computer simulation does not have internal validity issues. Even in non-deterministic simulations, Monte Carlo methods help in averaging out the influence of an independent variable and a dependent variable and showing if this influence is statistically significant. However, internal validity-like issues appear during the computer programming of a conceptual model into a computerized model (Sargent, 2005). In computer simulation literature the mitigation of this validity threat is done using verification activities. In gaming simulation sessions, the introduction of game players makes the experiment inherently open, allowing all sorts of confounding variables to distort the causal picture of one independent variable and one dependent variable. Furthermore as more soft variables are used to assess system behavior, e.g. work load and resilience, which do not need to be fully operationalized, gaming simulation, more than computer simulation runs the risk of not measuring exactly that what was intended to be measured.

The main objective in experiments in the analytical sciences is to manipulate on one or more factors (independent variables) and measure its effects on the manipulated variable (dependent variable) with a strong reliance on quantitative statistical methods (Zechmeister et al. 2001). The difference between experimental designs is related to the approach for which the sample procedure is conducted, whether a control group has been applied, and when and which measures have been used (see also Table 3.4) (Creswell 2003). Experimental designs are also known as a configuration of set of research design characteristics, e.g. a one-shot case study is a form of a pre-experimental design, which includes no random sampling, no control group and solely a posttest.

Traditional experimental research usually takes place in a laboratory setting, which is characterized by low contextual cues. Field experiments on the contrary are a type of experimental setting that pertain high contextual cues, in which often a representative sample of situations and participants are involved. (Harrison and List 2004; Vissers et al. 2001).

Table 3.4: Research design characteristics for three types of experimental designs.

Research design characteristic	Types of experimental designs		
	Pre	Quasi	True
Sample procedure	Non-random, e.g. convenience sampling	Non-random, e.g. convenience sampling	Random
Conditions	No control group	No control group/control group	Control group (and multiple group conditions)
Measures	Pretest and/or posttest	(Multiple) pretest(s) and posttest(s)	Pre and/or posttest

Harrison and List (2004: 1012) describe the difference between laboratory and field experiments by the following characteristics:

- Nature of the subject pool: the degree of a nonstandard, representative sample, e.g. professionals
- Nature of the information that the subject brings to the task: the field knowledge and expertise that the participants bring to the experiment
- Nature of the commodity: the presence of physical field characteristics in the experiment
- Nature of the task: the domain-specific tasks in the experiment
- Nature of the stakes: the urgencies of risks in field settings
- Nature of the environment that the subject operators in: the environment of the experiment

Based on these characteristics two more hybrid lab-field experimental settings can be identified. Artefactual field experiments relate closer to laboratory settings, to the extent that an abstract frame and imposed set of rules is used in combination with a higher degree of a representative sample of the researched population. Framed field experiments build on the characteristics of artefactual field experiments, but additionally entail the field context as well with regards to the commodity, task or information. Gaming simulation resembles mostly the latter type of field experiment, but differs by the use of game design components, which are the presence of facilitators, the use of game design principles and components, such as immersion and play and the emphasis on the value of the debriefing session.

All in all, laboratory and field experiments make a trade-off between internal and external validity by respectively guaranteeing that the treatment variable is the only variable impacting the experiment and by guaranteeing that the experiment provides enough contextual cues for the experimental results to also hold in real life. Since gaming simulation somewhat hovers between these two ends of a continuum, validity threats theoretically come from both sides. In addition, researchers use test methods like observations, surveys and interviews to see how the dependent variable reacts to the treatment. Thus, the external, internal and test validity need to be secured.

3.4.2.1 External validity

External validity is defined as "the extent to which findings from an experiment can be generalized to individuals, settings, and conditions beyond the scope of the specific experiment" (Zechmeister et al 2001: 161). Issues or threats that can occur for external validity are (Campbell and Stanley 2015: 6):

- Reactive effect: the effect of the pretest on the participants' sensitivity or responsiveness to the experimental variable
- Interaction effects: the interaction effects of biases in the selection of participants and the experimental variable
- Reactive effects of experimental arrangements: effects of the experimental variable upon participants being exposed to it in non-experimental settings. These include behavioral reactions of participants to the knowledge of being observed (e.g. Hawthorne effect) and the interactions between participants (contamination). When one of these validity threats occur in either one of the groups, but not in both, this becomes an issue for internal validity
- Multiple-treatment interference: effects of prior treatments remain present, thus possibly interacting with the new intervention

3.4.2.2 Internal validity

Internal validity is defined as the ability to confidentially "state that the independent variable caused differences between groups on the dependent variable" (Zechmeister et al 2001: 149). In order to make a causal inference, the experiment needs to establish a relationship between the independent and dependent variable, the cause must precedes the effect, and finally, plausible alternative explanations should be ruled out. To ensure the latter, the following factors (confounding factors or internal validity threats) need to be controlled (Campbell and Stanley 2015, Zechmeister et al. 2001):

- History: specific events that might occur between the first and second measurement next to the experimental variable
- Maturation: natural changes of participants over time, e.g. tiredness
- Testing: the effects of taking a test on subsequent testing
- Instrumentation: changes in the measurement of participants, due to the calibration of a measuring instrument or changes in the observers
- Regression: changes in the performance of participants that are due to the selection of participants on the basis of their extreme scores
- Subject mortality: loss of respondents in the different groups
- Selection: difference in individuals between the groups at the start of the study
- Interaction with selection (or selection-maturation interaction): different response of one group of participants to other internal validity threats, such as history, instrumentation

3.4.2.3 Test validity

Finally, an experiment needs research methods to extract the information about causality from the experimental run. In a computer simulation, the information is mainly about primary qualities, such as speed, travel time or punctuality. In gaming simulation often dependent variables, or constructs, come in the form of more secondary or subjective qualities such as work load, operator reasoning or quality of the handling of disruptions. This adds to the importance of measuring exactly what was intended to be measured. This test validity refers to the validity of measurement instruments, in which the following three types are in line with the American Psychological Association (Van den Brink and Mellenbergh 1998):

- Construct: the extent to which the instrument measures what it is supposed to measure
- Content: the extent to which the test can be reflected to a spectrum of situations or topics
- Criterion: the extent to which the test correlates to one or more external variables, which are a direct measure for the variable.

3.4.2.4 Robustness and reliability

Identifying the sensitivity of the experiment is strongly applied in computer simulation experiments, in which the researcher determines whether similar causal relations can be found when the experiment is repeated with exactly the same sample and setup. This complexity perspective on reliability follows from experiments with dynamic feedback systems. Because dynamic feedback systems inherit stochastic and (chaotic) properties, different results can be found when experiments are repeated with (almost) the same starting conditions. Indication of the sensitivity of an experiment is useful to assess whether the results are sensitive to the initial conditions or to game player's critical decisions.

3.4.3 Assessing gaming simulation from the design science perspective

The design-scientific model on the relation between gaming and policy making pointed to two relevant phenomena: firstly, game design and employment can occur relatively independent from the process in which it is embedded and secondly, effort is needed to translate gaming outcomes to this process. We feel that his conception of gaming and innovation resembles to a great extent notions found in the literature on sociotechnical transitions, such as the Multi-Level Perspective (Geels, 2002). In theoretical work on transitions, scholars pay much attention to the development of concurrent systems in niches (Rip & Kemp, 1998; Geels, 2002; Geels & Schot, 2007; Rotmans & Loorbach, 2009). Working as incubators, these niches allow for promising innovations to defy the negative selection environment that the current regime poses. From these notions we distill three relevant qualities of gaming simulation: 1) the relative difference, or innovativeness, of the concurrent system that a gaming simulation is able to envision and test, 2) the extent to which such envisioning can occur in a multi-stakeholder setting and 3) the extent to which gaming simulation enables these stakeholders to plan the transition process itself.

3.4.3.1 Search distance

The extent to which gaming simulation enables the envisioning of radically new innovations is termed the "search distance": the number of simultaneous changes in the significant elements of a railway system that can be realistically portrayed using this method. We add here that in multi-actor settings, looking for distant alternatives is often inhibited by the current regime. Search distance, therefore, is not only a technical aspect of a tool such as gaming simulation. By allowing for the creation of a transition arena, decision-makers are able to resist the technological paradigm that has kept them focusing on innovations that followed from the system itself. Such an arena would involve allowing several front-runners or change-inclined regime players to envision new directions for the system in these experimental and simulated niches (Rotmans & Loorbach, 2009).

Search is seen as a major part of organization problem solving and decision making (Cyert and March, 1963; March and Simon, 1958). As a field of inquiry, search has regained popularity since the introduction of the NK-model into the social sciences (Mihm et al., 2010). NK-models describe complex systems in an elegant manner by focusing on the amount of elements a system has (N) and the level of interrelatedness between these elements (K). A more elaborate overview of this framework can be found in Chapter 4. Design, and innovation, as a search activity can then be described as looking for different configurations of elements of a system (the design of an innovation) and subsequently altering the relevant elements from a base position towards this new configuration (the innovation process). In high-epistatic systems, i.e. systems with many interrelated elements or high values for K, the interaction effect of multiple elements causes multiple locally optimal hills to exist (see Figure 3.5). Thus, in these complex systems, optimization often involves measures in several elements simultaneously.

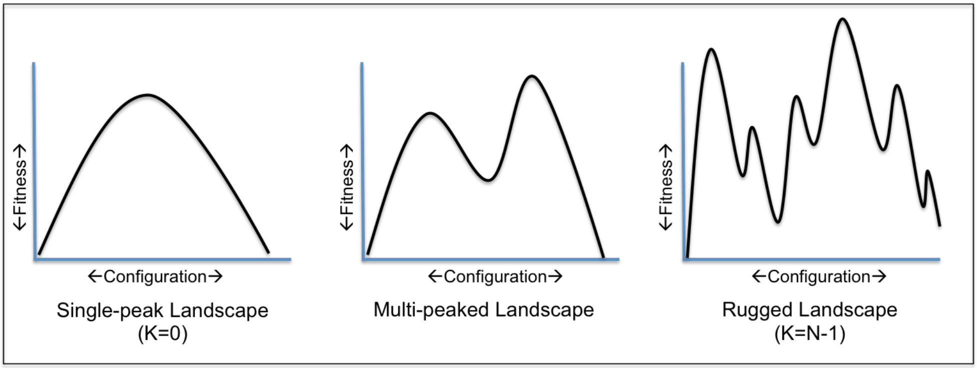


Fig. 3.5. Fitness Landscapes for simple and complex systems.

Within this set of elements, innovation actors search for different configurations that lead to an improvement of fitness (Levinthal, 1997; Frenken, 2006; Gavetti and Levinthal, 2000; Rivkin and Siggelkow, 2003). In general, search consists of two dimensions: depth and breadth (Gavetti et al., 2005; Katila & Ahuja, 2002; Laursen & Salter, 2006; Rosenkopf & Almeida, 2003). In terms of the NK-framework, depth would involve the amount of configurations under consideration of a searcher before choosing the most optimal one, while breadth would involve the distance these configurations can have to the configuration the system is currently in. Whereas depth relates to the carefulness by which alternatives are considered, breadth relates to the newness of these alternatives. These basic dimensions allowed us to characterize search using a simple 2x2 typology, which we already described in the theoretical chapter of this thesis. For clarity purposes we depict these in Table 3.5 and briefly reintroduce the four search strategies:

Table 3.5. Four search strategies

	High Search Depth	Low Search Depth
High Search Breadth	Exhaustive Search	Lucky shot search
Low Search Breadth	Greedy Search	Myopic Search

We note here that these strategies are not mutually exclusive and that firms may employ a mixture of different strategies, for instance comparable with Etzioni's mixed scanning approach (Etzioni, 1967) where broad design spaces are explored and a few promising directions are then more deeply investigated.

Exhaustive Search

The theoretically rational searcher would be able to find and evaluate all possible alternatives. Exhaustive search would thus involve studying all possible configurations in the design space and finding the optimal one. However, exhaustiveness is costly and time-consuming (Kauffman et al., 2003). For an individual searcher, the marginal gains from additional search will decrease as the already found solution is relatively fit, the basic argument for Simons' satisficing decision maker (Simon, 1959). Organizations as collections of searchers offer more possibilities for exhaustive search by allowing multiple searchers to search the technology landscape in parallel (Sommer and Loch, 2004), though technological regimes often cluster the search efforts in particular directions (Dosi, 1982; Knudsen and Levinthal, 2007; Kornish and Ulrich, 2011).

Greedy Search

Although a form of local search, greedy search still adheres to some aspects of rational decision-making: within the boundaries of myopia, all possible configurations are considered (Frenken, 2006). That is, from the current configuration of the system searchers explore all system configurations that differ by one element. Greedy search is a strategy found in cases where search is decentralized but coordinated through communication or hierarchical control. In both ways, incremental steps are only taken when the overall fitness of the system is increased.

Myopic Search

Myopic search is often related to decentralized and uncoordinated search. In this type of search, search responsibilities are distributed on an element basis. Thus, searchers only alter their element if the alteration causes the elements' contribution to the overall fitness value to increase, independent of its effects on other elements. Varieties of myopic search are extremal search (Frenken, 2006) where only the worst performing element is changed, or anchored search where elements are sequentially optimized according to the extent they perform lead functions (Mihm et al., 2010). In both cases, myopic search can lead to oscillations or 'chase their own tail' iterations (Mihm et al., 2003; Mihm et al., 2010): one element is optimized, but subsequent changes in other elements make this optimization obsolete. On the other hand, the lack of coordination also makes this search strategy less costly and time-consuming than the more extensive search strategies. Innovation research therefore has also focused on the changing the architecture of systems instead of its configuration (Henderson and Clark, 1990). Especially more modularized systems (thus lowering the value for K) allow for myopic search to become effective, since modules of elements can be optimized irrespective of other modules (Ethiraj and Levinthal, 2004; Frenken and Mendritzki, 2012)

Lucky shot Search

Rather than an intelligent approach, search can also involve looking for randomly chosen configurations far away from the current configuration. In this case, search breadth is high (many elements are changed) but depth is low (it only involves the evaluation of one configuration). This search strategy is often related to the discovery of radically novel solutions and the overcoming of lock-ins in local, but suboptimal, optima. However, its lack of depth hinders the learning effect of search. For more complex systems hold that configurations might be dramatically better or worse than its neighboring configurations due to the ruggedness of the technology landscape, severely limiting the robustness of the test results (Erat and Kavadias, 2008) and for a searcher it remains hard to determine what specific element or mechanism mainly determined the final fitness value (Loch et al., 2001). For systemic innovations, the high number of interrelated changes to elements that are needed is problematic for robustness: slight changes to the design of the innovation, either unplanned or the result of adaptation along the way, might render the outcome useless. For example, although removing 100 railway switches around a railway station might be beneficial, the ruggedness of the landscape impedes us from concluding anything about the value of removing 90 or 110 switches. Since a fitness landscape is never fully understood and knowledge about the landscape is distributed across multiple organizational entities, knowledge is highly contested. A current example is the debate about whether Japanese technical rail transport principles, to which the removal of switches belongs, can be copied in the Dutch network, and to what extent disregarding the cultural specifics of Japan might pose a problem. Thus, we see in this example and the example mentioned above that both the innovation itself as well as the information about it is contested.

3.4.3.2 Collaborative experimentation

Fitness values are not single indisputable metrics on the performance of a system. To begin with, in a multi-actor setting, different actors will value the system in different ways since they act on different incentives. Although decision-support tools will not align these incentives, they may help in determining to what extent the current regime will be replaced by the simulated niche system. For example, one aspect of making the railway system in the Netherlands more robust is the desire to remove railway switches. These elements make the system overly complex and interdependent, increasing the likelihood that initial delays will cause other trains to be delayed as well. Furthermore, the elements themselves are prone to breakdowns and failures. In this case, the infrastructure manager is incentivized to decrease the number of railway switches to allow for greater separation of transport corridors and a more reliable infrastructure. The train operating company, however, contests this measure for two reasons. Firstly, they expect an increase in transfers for passengers, as many connections will involve switching from one corridor to the other. Secondly, they fear a decrease in flexibility and expect fewer opportunities to maintain service levels in the case of a disruption.

For a concerted transformation, decisions need to be based on a shared set of understandings, increasing the need for a shared knowledge base. In these multi-actor settings, the contested nature of knowledge and misaligned incentives thus poses a challenge for decision-support tools. Far more than a matter of simple project management, the multi-actor setting of these transformations will need a process management approach that is able to cope with the inherent capriciousness. Thus, to overcome the problems of the contested nature of knowledge, decision-support tools should allow for the joint production of knowledge (Van Buuren & Edelenbos, 2004).

This joint production of knowledge or joint fact-finding requires that "stakeholders with differing viewpoints and interests work together to develop data and information, analyze facts and forecasts, develop common assumptions and informed opinion, and, finally, use the information they have developed to reach decisions together" (Ehrman & Stinson, 1999: 376). Since knowledge production is the goal of any research, the joint commissioning of research involves negotiating beforehand about the research questions, the demarcation, assumptions and methods (Koppenjan & Klijn, 2004). A process management approach prevents the exclusion of actors in the decision-making process (Termeer & Koppenjan, 1997). If excluded, these actors might contest the knowledge at a later stage or block the decision. Actors that are involved in the gathering and assessing of information on which an agreement is based are more inclined to support this agreement (Ehrman & Stinson, 1999).

Joint fact-finding is particularly important for transformations in networked infrastructures, since simulated niches (e.g. designs concocted in a gaming environment) are not self-replicating in such a context. A simulated experiment will not, through any mechanism, grow to replace the current system. Recall that networked infrastructures are inherently hostile to radical innovations. Consequently, the current system and its regime actors need to adopt the principles that follow from the simulated experiments and the design concocted there.

3.4.3.3 Transition planning

Whatever design follows from the interplay between the coalition of change-inclined regime players and the overall current regime, the path by which this design can be realized needs to be determined. Large projects such as these are often characterized by high degrees of equifinality: multiple ways can be found to reach the same end goal (Korsten et al., 1996; Alkemade et al., 2009). Multiple pathways exist between the current point on the fitness landscape and the desired point somewhere else on this landscape. Given that systemic transitions inherently involve a temporary decline in fitness (one traverses a valley to arrive at a different peak) it is worthwhile to consider where such decline is located and who of the stakeholders will mostly be hurt by such a decline. Furthermore, the systemic nature of the innovation will involve operational personnel changing the way they work. Frameworks in transition management have been criticized for their neglect of agency and for being overly functionalistic and teleological (Geels & Schot, 2007). Agency would involve the extent to which elements can be steered by a single decision-maker and which lie beyond their reach. Although removing a switch will not create much resistance, changing roles and rules for operators could very well lead to second-order effects, caused either by a lack of capacity or a lack of willingness to carry out the newly designed task. Knowledge about both aspects, the order of each step and agency, will increase the likelihood that a simulated niche design, in a gaming environment, will eventually replace the old one in real life.

In conclusion, perspectives on the use of gaming simulation from both the analytical and the design sciences delivered a range of performance criteria. Whereas the first perspective solely looks at gaming's experimental value via constructs such as internal and external validity, the design science approach added the notions of searching for different configurations of a system and gaming's ability to do so collectively. In addition, it also added the insight that gaming simulation might not merely be determining if an innovation works, but also through what transition process such innovation may be implemented. We add here that especially the design science approach needs a far more detailed understanding of the context-of-use and that other, more refined, criteria may come to light. That endeavor is exactly the core of the rest of this thesis. The criteria are shown together in Table 3.6.

Table 3.6: Overview of criteria for decision-support tools

Approach	Criteria	Sub-dimension	Mechanism
Analytical Science	Experimental value	Internal validity	Test 'true' value of innovation inside game
		External validity	Determine value of innovation in real-life context
		Reliability	Determine robustness of results against alterations in underlying parameters
Design Science	Niche building	Search breadth	Amount of runs available to test different sets of changes
		Search depth	Amount of changes to be tested simultaneously; ' newness' of the design
	Network building	Joint fact-finding	Mutual learning
		Fitness values of stakeholders	Determine how other actors value the system
	Transition planning	Determine agency	Determine feasibility of changes in elements
		Determine viable paths	Determine prerequisites for effective implementation of steps

3.5 Games as analytical science and design science instruments: an analysis

In the previous analysis of a range of innovation projects we found that decision makers were severely limited in their search space, i.e. the range of elements that they could manipulate and study in an experimental environment. This limitation was caused by the nature of the tests they had available, but also by their own inclination to keep an innovation process linear and sequential. Apart from such inclinations, this part of the chapter analyses how gaming simulations, as positioned between computer simulation and live-tests in their openness, score on the criteria put forward in Table 3.6. The following games are used for this analysis:

Table 3.7 Games used for initial analysis

Games	Goal
ETMET	Simulating traffic control procedures under a future timetable
NAU	Simulating traffic control procedures under a future infrastructure layout
BIJLMER	Simulating future traffic control procedures with existing infrastructure layouts and timetables
Platform Overnight Parking POP	Simulating different overnight parking procedures for trains at stations and shunting areas.
1st Phase	Simulating different (standardized and free form) traffic control procedures in the first minutes after larger disruptions

In the analysis for this part of the chapter, we see that in three instances gaming simulation allows designers to increase their search space by incorporating multiple conditions to the design. However, in all cases the amount of treatment variables remained one. For instance, when two conditions were used, one condition always remained unchanged for both the pre-test and the post-test.

Gaming simulation uses real-life operators as behavioral input for a simulation run. This advantage also poses a disadvantage as finding available operators has proven to be cumbersome in multiple instances. Railway traffic control is a 24/7 operation and operational staff is scheduled accordingly. A fully random sampling procedure was impossible since operator availability was the decisive factor determining the sample. Furthermore we have learned through the course of executing gaming simulations that more experienced operators are more suitable than less experienced ones. Firstly, they are better equipped for new and complex problems, such as dealing with disruptions in general and under conditions of new innovations specifically. Secondly, we have noticed that using a certain level of abstraction increases the need for game players to translate this abstraction. More experienced players seem better able to do so.

We firstly elaborate on analyzing the validity threats that gaming simulation in the railway sector face. In Table 3.8 we provide an overview for the five gaming simulations and how they scored on external and internal validity and reliability of the results.

Table 3.8 Validity threats per gaming simulation

	ETMET	NAU	BIJLMER	POP	1st PHASE
Ecological validity	More abstract but still all relevant information presented to players	More abstract but still all relevant information presented to players	High detail, small errors in context cues caused problems for immersion	More abstract but still all relevant information presented to players	High-tech-low-tech-hybrid
Immersion	High	High	Medium	Medium - High	Low - Medium
Reactive effect	N.a.	N.a.	N.a.	N.a.	N.a.
Interaction effects	Unknown	Unknown	Unknown	Unknown	Unknown
Reactive effects of Experiment	Many observers; game players separated	Many observers; game players in one room	Many observers; game players separated	Low amount of observers; game players in one room	Many managerial observers; game players separated
Multiple-treatment	N.a.	N.a.	N.a.	N.a.	N.a.
Internal validity					
History	N.a.	N.a.	N.a.	N.a.	Learning effect for facilitators
Maturation	N.a.	During post-test some traffic controllers became tired	N.a.	N.a.	During post-test some traffic controllers became tired
Testing	Medium learning effect for traffic controllers	Medium learning effect for traffic controllers	High learning effect for traffic controllers	High learning effect for cleaning personnel	Intensive discussion about game and scenario between pre- and post-test
Instrumentation	N.a.	N.a.	N.a.	N.a.	Some observers were replaced during the experiment
Regression	N.a.	N.a.	N.a.	N.a.	N.a.
Subject mortality	N.a.	N.a.	N.a.	N.a.	N.a.
Selection	Respondents more experienced than their average real-life counterparts	Respondents more experienced than their average real-life counterparts	Respondents more experienced than their average real-life counterparts	Respondents more experienced than their average real-life counterparts	Respondents more experienced than their average real-life counterparts
Interaction with Selection	N.a.	N.a.	N.a.	N.a.	N.a.
Test valid.					
Construct validity	Resilience, Robustness	Resilience, Robustness	Robustness	Throughput capacity	Resilience, work load
Content validity	Punctuality and capacity as proxy, measured on train level	Punctuality and capacity as proxy, measured on train level	Punctuality and capacity as proxy, measured on train level	Amount of trains	Punctuality and capacity as proxy, Work load measured using self-rating
Criterion validity	Video, quantitative data, observers, debriefing	Quantitative data, observers, debriefing	Quantitative data, observers, debriefing	Quantitative data, debriefing	Video, questionnaires, observers, debriefing:

3.5.1 External validity threats

As we have stated, external validity issues appear when a design needs to be tested in a simulated experimental environment. Thus, building this environment needs to incorporate and preferably tackle these issues. We see some profound issues here that need further explaining. Firstly, models are inherently more abstract than the reference system. There seems to be a negative parabolic relation between abstraction level and ecological validity.

The BIJLMER game used little abstraction, but was deemed less realistic by traffic controllers due to slight changes in the interface. Other games were more abstract, e.g. using sponges for trains instead of the standard traffic control interfaces. By being more abstract, these models were less confusing to the game players and we saw in the debriefing that psychological reality was still perceived as high by them. Secondly, experimental arrangements might threaten the external validity. As far as we can see, two factors are most important here. Firstly, the benefit of gaming simulation is that processes that are normally spatially and temporally dispersed are now brought together. For designers, managers and decision-makers this allows them to study this processes in more detail. However, as they are observers that bring more scrutiny to the behavior of traffic controllers, strong Hawthorne effects might take place. Although inconclusive to this respect we see that each game caused high levels of immersion of the game players and we feel that this somehow decreases potential Hawthorne effects. This points to the first inherent dilemma of gaming in innovation processes: validity and usability are sometimes at odds with each other. However, both a highly observed game and a less observed game have both been validated in real-life and concerning potential effects of increased scrutiny on external validity we saw no indication.

3.5.2 Internal validity threats

Internal validity issues appear when other variables within the experiment might explain the change in the dependent variable as well. Since all gaming simulations did not use a control group, it is hard to control for internal validity issues. A critical examination of possible confounding variables in the five cases showed that learning effects of players and facilitators, player fatigue and dynamic instrumentation are the main factors decreasing the internal validity.

When using less experienced operators we see that high learning effects take place during a simulation run, making it difficult to compare a pre-test with a post-test. For instance, during the overnight parking game cleaning personnel had difficulties in dealing with the abstraction and game mechanics during the first parts of the gaming simulation. Additionally, during the 1st PHASE game we saw that game facilitators had problems in facilitating the game and became more apt only as the game evolved. Although in this case only a minor problem, it points to the importance of training facilitators in the task they are responsible for during the session. If neglected, the learning effect of a facilitator might be mistaken for a treatment effect. Finally, we have noticed how gaming simulation sessions are demanding sessions that drain the energy of game players. To still be able to realistically compare a pre-test and a post-test, experimenters should incorporate fatigue effects.

3.5.3 Test validity threats

Construct validity is hampered by a problem of focus and a lack of a clear definition of often used concepts like resilience and robustness. Furthermore, we saw in the NAU game that disagreements occur on what the focal construct should be. While ProRail was interested in system performance, the Dutch railways was more interested in what the effect of the innovation was on the work load of their train controllers. Related to content validity, we see a very narrow focus on resilience and robustness as the extent to which punctuality and capacity can be maintained throughout a disruption and that these proxies were measured on a train level and not on traffic level. However, this specific focus is also present in the reference system. A main and unique advantage of gaming simulation is that it easily allows for triangulation of data sources and thus increases criterion validity. In almost all instances we see that logs of punctuality are combined with observer logs, video reports and data from the debriefing to see if these data sources corroborate each other's findings.

3.5.4 Input for next cycle

In three cases the hypothesis was rejected (see Table 3.9), much to the surprise of the involved project managers. However, they saw the gaming simulation session as externally valid enough to trust the outcomes and included the findings in the continued work on their proposed solution. In addition, the gaming simulation gave much valuable and rich information about what measures where needed parallel to their solution. These measures could stem from the simulation sessions itself (endogenous) or could be signaled by game players during the run or after the session (exogenous). For instance, using time slots in controlling high frequency traffic did not work quite as expected, but game players signaled additional directions for improvement, e.g. by changing platform lengths and building a railway track dedicated for overhauling.

Table 3.9: Results from gaming simulation sessions.

Input for next cycle	ETMET	NAU	BIJLMER	POP	1st phase
Hypothesis	Rejected	Accepted	Rejected	Accepted	Rejected
Additional data	-	Validation; additional endogenous dominant parameters found: cooperation between traffic control echelons	Additional exogenous and endogenous dominant parameters found: infrastructure and procedural changes	Validation	Additional endogenous dominant parameters found

NAU serves as a prime example of a gaming simulation of which the findings could be to some extent validated in real life. Some months after the session, this new way of handling traffic around the central node of the Dutch network was indeed altered. Different from the game, the switches were kept and their nonuse could only be guaranteed by work arrangements. It appeared that the same behavior was seen in real life as in the game: stability of single corridors, e.g. Amsterdam - Den Bosch, was sacrificed for the robustness of the total network. However, the fact that the railway switches could still be used, mostly in situations where flexibility was demanded by traffic controllers, meant that the system had a natural tendency towards a less robust but more flexible way of controlling traffic. In 2015 and 2016, the measures tested in the NAU game were made permanent by changing the whole infrastructure around Utrecht station, decreasing the amount of switches five- to tenfold.

3.5.5 Niche building

In addition to being a valid experiment, our theoretical frameworks also led us to look at the extent to which gaming experiments allow for certain classifications of search, search that is only possible because the restricting forces of socio-institutional pressures are to a lesser extent present during a gaming simulation. For this purpose we compared two of the gaming simulations (NAU and BIJLMER), two games conducted at the beginning of this research project, with two innovation projects that involved solely a real-life trial. The latter projects have already been described in the first empirical analysis of innovation projects in the first part of this chapter, their experiments we name COUNTDOWN and PASSENGERS. We show the results in Table 3.10.

Project managers in all studied projects were restricted in their search space. Certain elements perform lead functions (Mihm et al., 2010): these functions are optimized prior to the problem formulation for the project managers. Planning a railway system is in many ways hierarchical, with infrastructure, safety and rolling stock elements being optimized before operational measures are considered (Goverde, 2005). In all four cases we see empirical evidence for this.

In COUNTDOWN we see that project managers simply accept this restriction and find a solution within this space. The signal box that was placed in front of the train did not interfere with any of the other elements such as IT and safety infrastructures. By doing so, the design team cancelled out potentially better solutions just to come to a proof-of-concept quickly. In PASSENGERS we see that the solution is also sought in this confined search space but that based on this solution, additional demands are placed on other elements. Given infrastructural, timetable and rolling stock constraints they found a solution for which the proper working could only be guaranteed if rolling stock composition and timetables were static. Because in gaming simulation the simulated system was easily modified, elements not considered directly under the influence of the project manager entered the search space. For instance, in testing traffic control based on corridors instead of geographical areas around Utrecht, the possible removal of obsolete railway switches was discussed and the BIJLMER game was carried out using a new timetable.

Table 3.10 Projects' search space when using live-tests and gaming simulation

	COUNTDOWN	PASSENGERS	NAU	BIJLMER
Goal	Decrease dwelling time variation at stations	Decrease dwelling time at stations	Make central node of network more robust	Find new control concept
Design	Design team found that delay in giving right-of-way an departure procedure caused dwelling times to be high and variable. Solution was to tell train drivers when to depart through a machine placed on the platform	Design team found that an uneven spread of passengers was the main problem. Solution was to provide information about train length on the timetable with colors corresponding with places on the station platform	High dependence between train services was the main problem. Solution was to control traffic according to corridors, separating railway lines	Traffic control based on time slots rather than bases on fixed time points. Concept needed for coping with higher frequencies of trains on main corridors
Design restrictions	No interference with infrastructure, safety and IT-systems	No changes in infrastructure and railway stations	Not explicitly mentioned	Only focus on operational measures
Elements changed	Additional signal on the platform	Timetable, information on platform, rolling stock composition	Work division between controllers, use of railway switches, operational roles and rules	Timetable, operational rules, communication
Build	Proof-of-concept built and tested on small-size station	Proof-of-concept built and tested on small-size station	Gaming simulation built in few weeks	Gaming simulation built in few weeks
Run	Test was carried out during other test involving increase in frequency. Heightened attention from management because of other test	Test was carried out under ideal circumstances (good weather, no last minute timetable changes)	Isolated run. Both large and small-scale disruptions were tested on their influence on robustness of the designed solution	Isolated run. Players had problems with unfamiliar interface of game. During run players pointed to additional measures needed to make the design work
Analyze	Dwelling times decreased but no evidence for a relation with the measure. Uncertain what the influence of managerial attention was on train driver behavior. Measure also impacted behavior of waiting passengers, who used the countdown machine as well.	Valuable measure but very low external validity due to many restrictive assumptions in the experiment. Test assumed static train composition and no last minute changes in platform assignments	Valuable, but additional rules for using switches needed as well as better cooperation between different echelons of train and traffic control	Valuable but additional measures needed regarding operational communication, infrastructure and stations.
Follow-up	New test planned involving stations and level crossings to synchronize departure procedure	Search for newer solutions that are less dependent on train composition	Solution implemented, renewed focus on cooperation between echelons	Solution not implemented, contingent on larger projects being initiated first

In both projects that used real-world case studies, the additional elements in the search space delayed the process. Only when other project managers changed those elements, the test could take place. In COUNTDOWN, ProRail needed cooperation of Dutch Railways for interfering in the operation of train drivers. Furthermore, they had to wait for the high-frequency train transport test to take place. In PASSENGERS, the design team looked for a specific day the train operating company could guarantee a static rolling stock composition and chose a specific railway station at which last minute platform assignment changes were not to be expected. Building a gaming simulation takes less time because manipulation of elements in the simulated system did not depend on the actions of other departments.

The run of a real-world case study is rigid. No changes could be made to elements during the experiment. During the run of a gaming simulation, elements that are part of the search space can be changed. For instance, the strictness of the separation between corridors was dynamically changed to test whether the separation would have to be made permanent and whether railway switches would become obsolete. By doing so, project managers were able to compare different configurations neighboring the configuration that was initially tested. However, such measures on-the-fly seriously impact the validity of a gaming simulation if purely looked at from an analytical-scientific perspective. This provides us with the second instantiation of the dilemma between validity and usability of gaming simulations.

Search depth is further increased with gaming simulation because game players are able to pinpoint where additional measures are to be taken if the solution needs to be further improved. During and after the BIJLMER game, traffic controllers and train engineers showed exactly where extra measures were needed if high-frequency train transport was to be realized using the new traffic control concept. Platform length and additional side-tracks were seen as critical. Project managers neglected these measures since the responsibility for these was located elsewhere. During evaluations of gaming sessions, project managers acknowledged that this information exchange between operational personnel and higher echelon decision makers is rare but valuable. We saw that designers and operators of the system speak different languages from highly conceptual to operational. Gaming simulation promises to serve as an intermediate level of abstraction thus allowing communication. Our findings seem to show how this communication can lead to greedier search. Search breadth within gaming simulation seems to have limits. The BIJLMER game signaled that the simulated system needed to resemble to a large extent the system the operators are familiar with. It points to two potential shortcomings. Firstly, as gaming simulation is a model of a real system, the more encompassing this reference system is, the more iconic certain elements of the simulated system will be. Especially the analogue versions of gaming simulation can only allow for a certain level of detail. The BIJLMER game, but also other gaming simulations, showed that this level of detail is critical for operators to understand the game. Secondly, studies on operator decision-making have shown how decisions are more based on if-then recipes than on a rational process (Greitzer et al., 2010). Higher search breadths will cause if-then recipes to become invalid. These recipes are built through experience and these might only hold for this specific system. Radical innovations therefore might be difficult to test with gaming simulation involving human operators that apply heuristics, if looked from an analytical-scientific perspective. This is the third instantiation of the dilemma between validity and usability. In this case gaming simulation may be well suited to envision and design radical innovations, but not able to validly test them.

Since in a case study less manipulation of elements is possible, project managers found it hard to understand what in the end caused changes in system behavior. In COUNTDOWN, managers saw a decline in the variety of dwelling times but were unsure if this effect was caused by their solution or by extra managerial attention. In PASSENGERS, they tested the solution under strict assumptions, which were only met on a specific day. After the test it remained highly uncertain if their solution would be valuable under less stringent assumptions such as a changing rolling stock composition and more last minute changes in platform arrival. The fact that gaming simulation allows for search to take place even during the run increased the confidence of the project managers in their solution. Results showed what additional measures were needed for their solution. Furthermore, as they were able to observe the system holistically (the games were played either in one room or in several adjacent rooms), they were better able to see what processes were mediator variables between their solution and the behavior of the system.

All projects started with a form of lucky shot search. They formulated a solution that differed from the current system in multiple elements and only this solution was tested. We note here that gaming simulation allows more elements to be changed without the interference of other departments and organizations, increasing the ability to test out more new systems. For real-world case studies this strategy remains in effect during the test. Little can be learned from the results and subsequent tests involve another lucky shot in a different direction. Gaming simulation allows for a greedy search to take place around the configuration that is tested in this lucky shot search strategy. Project managers are able to find out if in the vicinity of this solution more optimal solutions lie. Gaming simulation allows for a stronger feedback between test results and subsequent designs.

3.5.6 Joint fact-finding

In addition to the methodological functionality of experimenting using gaming simulation and the practicality of testing radically new innovations, we have stated that the social reality of innovation processes impacts the value of gaming simulation. The extent to which multiple stakeholders can, with the support of gaming, achieve progress in the innovation process then becomes important.

The closed character of computer simulation creates problems for the joint commissioning of research if compared to its more open variant of gaming simulation. Since simulation results are highly susceptible to the validity of the underlying assumptions, as well as to initial conditions of the model (Frigg & Reiss, 2009), network actors or external stakeholders can contest the outcomes. Convincing other actors to form a coalition becomes problematic. As those involved in computer simulation within ProRail have noted, it is not uncommon for simulation results to end up being shelved. In this regard, gaming simulation has a significant advantage. Since multiple actors are involved in making, executing and evaluating the game, gaming simulation better allows for joint fact-finding. Two cases serve as an example here.

To test the feasibility of different train control concepts in handling larger disruptions under conditions of high frequency passenger rail, a gaming simulation session, ETMET, was organized in which traffic and train controllers had to manage a disruption on the corridor between Amsterdam and Eindhoven. Although at first the measure was purely an initiative of ProRail, the involvement of Dutch Railways in the test led to the mutual understanding that the traffic and train control processes were highly interlinked and an optimization of these processes required the concerted effort of both parties. Furthermore, it proved that a concept whereby controllers only have to adhere to a certain pattern of disruption management rather than a fully prescribed way of handling disruptions would not lead to better results. This mutual learning could only be achieved through the concerted effort of both parties to instigate and execute the required research. Far more than computer simulation, gaming simulation allows for this to happen.

In another case, the game NAU, in which project managers assessed the effects of assigning traffic controllers to certain corridors around the station of Utrecht on network performance, the interplay of different actors led to better insight into the motivations of each actor involved. Although for ProRail the strict separation of corridors was believed to be a means by which a more robust traffic flow could be guaranteed, for Dutch Railways the test was about lowering the work pressure experienced by their train traffic controllers. Better insight was only achieved through the cooperation of these actors during the gaming simulation design and employment phase. However, this dialogue was not inherent to the more analytical-scientific approach initially taken to design the game. Had this dialogue not occurred, the differences in motivations would potentially have caused the decision-makers to make different decisions based on the same (perceived as valid) information. Clearly, this would be detrimental to a concerted effort to transform the system and points to the limited value of valid knowledge in such multi-actor environments.

If we could summarize the basic difference between computer simulation and gaming simulation it would be to describe gaming as opening the black box of simulation. In fact, decision- and policy-makers, designers and managers are able to experience the simulation taking place, being amongst the game players and observing them interact closely with other players. Powerful results not only stem from quantitative evaluation of system performance, for example, by counting delays, but also from qualitative assessments by both observers and game players. We believe that this is where learning from each other occurs and a shared understanding of the intricate complexities of a sociotechnical system arises. In several instance have we noted that such results are at odds with each other: analytical-scientific and design-scientific ways of embedding gaming simulation in innovation processes are sometimes mutually exclusive.

Although a somewhat neglected activity (Crookall, 2010), proper debriefing then becomes highly essential, since this is where qualitative judgments about system performance can be shared and differences in perspectives can lead to synergistic results. Often we reserve an hour for such a debriefing, where we lead the discussion as a neutral moderator in a forum setup. It is in the aftermath of a highly intense gaming session that individual experiences, differences in insights and general conclusions become apparent. Although still underdeveloped, a debriefing session appears to be crucial to retrieve valuable information, especially when it involves rich qualitative data. Chapter 9 will therefore provide a more thorough elaboration on this part of gaming simulation.

Additionally, although computer simulation allows for multiple runs to test the robustness of the results and its sensitivity to starting parameters, gaming simulation usually encompasses a handful of runs due to time constraints. This raises another important reason for a high-quality debriefing. In such a forum setup, game players, observers and facilitators can discuss to what extent system behavior was sensitive to a few critical game player decisions and what validity threats this sensitivity might pose. Future research will therefore focus on how to make qualitative assessments during a game, how to synthesize these in a debriefing and what strategies we can use to collectively assess the robustness, usability and credibility of gaming simulation outcomes.

3.5.7 Transition planning

In the gaming simulation about traffic control concepts around Utrecht, the poorly defined decision responsibilities of the local traffic controllers and the national traffic controllers respectively, became apparent: this problem inhibited the effective implementation of the new traffic control concept. As for transition planning, gaming simulation is very suitable for determining the specific sequence of steps that would underlie these transformations. In this case, the organization needed a better definition of what each level of traffic control would be responsible for before it could take the step of adjusting the traffic control concepts. In a similar vein, a game involving the testing of traffic control concepts around Bijlmer Junction provided insight into the value of these concepts, as well as the prerequisites required for such a measure to be valuable. Before really implementing this measure, ProRail found that additional measures such as extra buffer times and more space for the sidetracking of trains were essential. Both examples again emphasize the importance of a debriefing. The additional information, such as the additional infrastructure needed, does not automatically flow from the game. In both instances, these qualitative findings were retrieved in a debriefing session.

Since simulations are basically large-scale computerized thought experiments based on a set of assumptions, such tools will not be able to tell us anything about the extent to which elements can be designed and thus controlled. As agency is especially important in sociotechnical systems, the transfer of system configurations from experimental simulated niches to the actual current regime is not a matter of simply copying them. Social actors within these systems often have relative autonomy, and their adherence to prescribed roles and rules is not guaranteed. In multiple cases, gaming simulation has shown decision-makers the extent to which the technical and social solutions would be adopted by other actors in the network or by operational personnel. For instance, the VECHTBRUG Game, which tested the feasibility of traffic controllers allowing for the opening of a drawbridge in a regime of high frequency transport, showed that gaming was a valuable tool in testing for agency. Thus, although adding gaming elements to a simulation decreases the search distance, it does reveal those system configurations that can feasibly be changed. Computer simulation shows what could be achieved theoretically through process innovations, gaming simulation shows what can be achieved practically, given the limitations to the control of the elements by decision-makers.

3.5.8 Conclusion on the theoretical value of gaming simulation

We have looked at the value of gaming simulation as a pure experiment and in relation to real life tests and computer simulation when looked at its ability to allow for specific search strategies, joint fact-finding and transition planning. The introduction of game players instead of assumptions into the model increases the external validity but this seems to come to the detriment of internal validity. Humans, as players, facilitators and observers make the method inherently more messy and this demands from the gaming professional to better manage the exercise. However, the introduction of humans also allows the method to support joint fact-finding, more so then more closed exercises such as computer simulation. With regards to search, not merely looking at the extent to which the experiment is valid but also at the extent to which multiple parameters can be manipulated at the same time, we see that gaming is especially well equipped to enlarge the search depth, rather than the search breadth. Computer simulation, with humans absent in the model, is better able to test innovations more radically different from the current setup of the system. With gaming's larger search depth however, we see that games are better able to test the robustness of innovations to changes in the makeup of the innovation itself as well as provide more information about the specific planning of the transition towards a newly desired end-state through smaller steps.

3.6 Conclusion

This chapter provided a first overview on how innovation processes take shape in reality and the role two different models play in this manifestation. Additionally, the second part looked at gaming from these two models and concluded that validity and usability concerns, related to two different approaches, are at odds with each other. A valid game may not be usable, a usable game may be invalid. However both the analyses focused on incremental innovation processes. Given the relation between games and innovation processes (DIS vs DIL), and the fact that systemic innovation processes may make other qualities of gaming ‘ usable’, the conclusions from this chapter have a limited reach.

3.6.1 DIL in systemic innovation processes

If we compare the studied incremental innovation projects with the envisioned, and more systemic, projects currently undertaken or planned in the near future, we see the dilemma facing the sector in Figure 3.6:

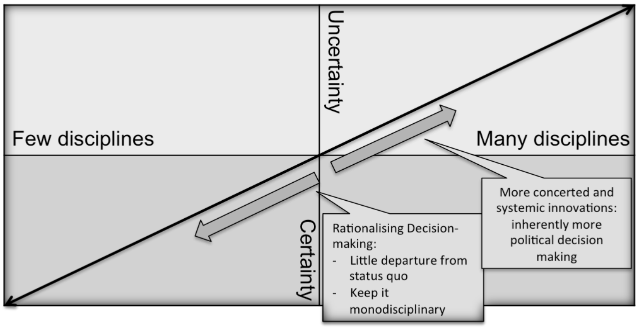


Figure 3.6: From rational decision making towards political decision making

So we state that the more systemic process innovations of Robuust Spoor, PHS or Mistral , mentioned in the previous chapter, inherently cause the decision-making processes to be more political. Two arguments can be given: firstly, the substantive uncertainty is higher since measures are new and untested. On top of that, systemic innovations are programs of subprojects that are highly interdependent and need a concerted effort of multiple parties. Furthermore the fact that the systemic innovations tackle the complete system makes it oblivious for departmental and organizational boundaries. Where decision makers in the studied innovation projects refrained from too systemic innovations, programs like ‘robuust spoor’ and ‘kort volgen’ need the effort of the infrastructure manager (adjusting infrastructure, safety systems and traffic control concepts), train operating companies (adjust rolling stock, training and scheduling of personnel, adjust train schedules), municipalities and ministries. All these parties have different goals and strategies within the railway sector which conflict and the usual ways of coping with this (making it incremental) are both not desirable and feasible.

Since railway systems are technically and socially complex, rational, sequential and linear models on decision making and innovation cannot explain the innovation processes found within the railway sector if they involve systemic change. First of all, the technical complexity makes it impossible for decision makers to calculate the consequences of all systemic alternatives: they are uncovered 'on-the-fly'. Secondly, within a socially complex environment, stakeholders are highly reliant on other stakeholders with differing values, goals and incentives. Especially in the railway sector, since the unbundling of vertically integrated network-based industries, possibilities for direct steering by one actor for innovation purposes have diminished (Weijnen and Boumans, 2006) and "hierarchy as an organizational principle has lost much of its meaning" (Koppenjan and Klijn, 2004: 3). Therefore, to study processes of systemic innovations in railway systems, simpler analytical-scientific and reductionist models therefore seem less suitable. Linear innovation models might help to explain more incremental innovation processes but when innovation processes start to involve systemic change, these models will inherently fall short. Additionally, the theoretical frameworks have to account for the fact that innovations are not atomistic entities but embrace a multitude of elements and hence creates dynamics in the innovation *artifact* itself, not only in the social spheres of the innovation *process*. Theoretical frameworks put forward in chapter 4 therefore need to account for these characterizations. In addition, we argue that decision making on systemic innovations in railway systems takes place in a network of independent actors and more non-linear perspectives are better suited to study processes in this context.

3.6.2 The need for a design science approach

This first empirical chapter looked at via what models initially innovation stakeholders conceptualized the innovation process and how particular decisions made by these stakeholders showed how such models became normative in real life. Over a range of incremental innovations, the process was intendedly designed as a linear and sequential process. Stakeholders did so to be able to quickly test a proof-of-concept in an experimental setting. This shows that potentially the use of gaming as pure experiments forces innovation stakeholders to make dysfunctional design decisions, especially for systemic innovation processes. On the other hand this chapter also showed that simply moving from an analytical science to a design science approach is no trivial matter. Whereas from an analytical science perspective designing a gaming simulation can take into account relatively simple performance measures, borrowing from the literature on the design of classical experiments, using the design science perspective demands a thorough analysis of the context-in-use. This chapter made a first step, using a simple and rudimentary coalescence of theoretical notions from the literature on sociotechnical transitions and complex systems. However a more detailed description and understanding of the true context-of-use will determine what makes a gaming simulation in a particular context, and in a particular point in time, usable. The following chapters, both theoretical, methodological and empirical, precisely constitute this analysis of the context-of-use.

In addition, an important factor in explaining the halting of innovation processes in the cases in this chapter was the perceived low validity of the results of the experiments and the resulting uncertainties when designs from such experiments would be implemented in real life. This points to the fact that validity concerns play a large rol in translating gaming outcomes to design-in-the-large processes, whatever the usability of these outcomes may be. Hence, when taking a design-scientific approach to gaming simulation we still acknowledge the relevance of validity for the context-of-use. We posit that this relates strongly to the credibility quality described in a design-scientific approach to game design (see Barreteau et al., 2001; Klabbers, 2009).

4 Theoretical framework

We start with the assumption, resulting from the previous chapter, that the complexity of innovation processes is highly influenced by the complexity of the artifact to be innovated and the complexity of the context in which this artifact operates. Also we assume that the dynamics in these two spheres highly interact. In the previous chapters we termed these technical and social complexity but we add that theoretical frameworks might lead to more detailed and fine-grained classifications. Especially, we state here that the capriciousness of the innovation process, or its highly nonlinear nature, is a consequence of the many dimensions of complexity that characterize railway systems. Railway systems, as we have demonstrated, are complex sociotechnical systems. Hence, its evolution is determined by technological, social, institutional and political factors (Schot, 1992; Künneke et al., 2010). Theories on innovation and innovation processes should therefore be closely linked to the specifics of the artifact that is to be innovated and the environment in which this happens. Before exploring the many perspectives on innovation that emerged over the decades, we provide a more detailed characterization of railway systems. This helps in understanding why certain theoretical frameworks are applicable and why some are not. We will point to its complex nature, its networked properties as well as the dynamic interplay of technical and social elements over time as the defining qualities of railway systems. We briefly explore a few notions that emerged from different disciplines studying networked infrastructure in more general terms that help in defining railway systems.

4.1 Systemic perspective

A systems, or complex systems perspective has pervaded throughout many different fields that have studied transport networks. This has resulted in a plethora of different characterizations of these networked infrastructures. Therefore railway systems can be called communicative infrasystems (Jonsson, 2005), sociotechnical systems (Geels, 2002, Markard, 2008), systems-of-systems (De Neufville and Scholtes, 2011; De Weck et al., 2011), complex adaptive systems (Holland, 1992), lambda-systems (Nikolic and Dijkema, 2010), open-assembled systems (Tushman and Rosenkopf, 1992) or large technical systems (Hughes, 1987). Although all characterizations slightly differ on where they put emphasis on, in general a set of notions holds for every perspective. That is, railway systems show systemic and networked properties, are complex bundles of co-evolving technical and social elements and serve a set of different public values (Thissen and Herder, 2003), hence its evolution over time is complex and intractable (Nikolic and Dijkema, 2010). Many of the different perspectives acknowledge to a certain extent the same set of stylized facts. Therefore the theoretical framework will not be written from one specific theoretical background but will employ theories, frameworks and nomenclature from a plethora of perspectives.

Firstly, railway systems are systems. This seems trivial but is highly valuable for a systems perspective gives way to analyzing new properties that one would not observe if solely looking at railway networks as atomistic products. Systems are a set of components that serve a common purpose (Ackoff, 1979). In its most abstract form, systems consist of components, linkages that connect components and interfaces that allow for communication between components (Tushman and Rosenkopf, 1992). The fact that they serve a common purpose already hints to complexity as a defining characteristic of systems. Multiple components are interdependent to varying degrees in serving the common functionality of the system. Irrespective of degree, interdependence means that optimizing such a system is a non-trivial activity for optimal changes on a component level may be suboptimal on a system level (Simon, 1969; Kauffman, 1993). Changing a component or replacing a linkage may cause second-order effects because the change affects the functionality of other interdependent components. That is, there is a nonlinear relationship between the structural configuration of the system and its functional properties. For example, optimizing a timetable may theoretically increase the functionality of the railway network. However, it also may place additional burdens on already overloaded traffic controllers, diminishing the initially expected functionality increase. Systems thus are to a certain extent complex in that it is difficult to predict the properties of the system even if the properties of the system elements are fully known (Wiener, 1948).

4.1.1 Networked properties

These systems are large in scale (Gokalp, 1992) and display networked properties. That is, in explaining how a system functions it is not merely the composition of elements one should look at but also the topology by which subsets of elements are spatially located. For instance, the challenges facing the Dutch railway system are of a different kind then the challenges facing the UK railway system, because of its highly polycentric nature. In our example mentioned above, optimizing a timetable for one part of the network might create overall sub-optimality given the timetables for other parts of the network. This adds to the complexity since it is not only the component structure of the system but also the spatial structure of the system that defines the overall functionality. Furthermore, these systems tend to converge with other systems to form evermore multi-technological networks (Thissen and Herder, 2003). In railway systems we see for instance the convergence of transportation, communication and IT-networks.

4.1.2 Sociotechnical properties

The notion that most systems are not merely collection of technical elements rose out of the research done at the Tavistock Institute on work systems in UK coal mines (Trist and Bamford, 1951; Emery, 1959; Trist, 1981). This notion answered to a technocratic view on work systems prevalent in the scientific management movement (Taylor, 1911). In the latter movement, complexity was deemed too high for individual workers to cope with demanding from organizations to break down work processes in ever increasing specialist tasks. The sociotechnical systems theories holds that optimizing work systems is best done by joint optimizing the technical part and the social part, a departure from a Taylorist focus on optimizing solely the technical part. On a broader scale, transport systems are sociotechnical systems as well (Wilson et al., 1997; Geels, 2004; Markard, 2006). A sociotechnical system is a "purposeful system that is open to influences from, and in turn influences, the environment (technical, social, economic, demographic, political, legal, etc.); the people within it must collaborate to make it work properly; and success in implementation of change and in its operation depends upon as near as possible jointly optimizing its technical, social, and economic factors" (Wilson et al., 2007: 102). A modern day railway system comprises of railway tracks, signaling systems, railway switches, trains and many more aspects. As such, the system can be seen as a technical system. However, for the working of a railway system, mere technical equipment does not suffice. Traffic dispatchers, train engineers, maintenance personnel are needed to make trains run on the network and institutions are instigated to organize the work in a safe and economically efficient manner. In a railway system therefore many different systems with differing 'openness' jointly work together to serve a common purpose. Technical artifacts, humans and social organization are three elements found in railway systems and studying these systems therefore has to take into account the different rationalities these subsystems have.

According to Geels (2005) in any sociotechnical system, three elements can be distinguished: the sociotechnical system itself, human actors and groups and rules and institutions. In case of the railway sector we could see the operation of the railways as a sociotechnical system with train engineers, traffic controllers, high-level decision makers as human actors organized in groups such as the Dutch railways, the government and ProRail. The cooperative arrangements, implicit and explicit, are coded through rules and institutions, which can be either formal or informal. In his model of sociotechnical systems Geels (2004) explicitly mentions the bidirectional influence these components have on each other (see Figure 4.1)

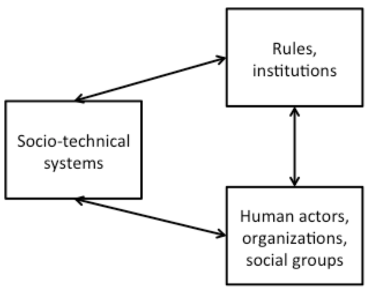


Figure 4.1. Sociotechnical systems and its components (adapted from Geels, 2004)

Sociotechnical systems are created, maintained and operated by human actors for without human actors, sociotechnical systems would not function. However, sociotechnical systems also influence human behavior and how this behavior is organized through rules and institutions. For example, the invention of the car meant we could travel with much higher speeds than before. Rules were adapted (think of speed limits) and behavior changed (people were able to travel longer distances, commuting distances grew). These bidirectional influences cause the components of a sociotechnical system to coevolve (Geels, 2004). Over time actors, rules and the system are configured to coherently guarantee a well-functioning of the sociotechnical system.

Because of this co-evolution, sociotechnical systems are characterized by a high degree of path dependence (Geels, 2004; Markard, 2006). Path dependence is the occurrence of one specific contingent event having a persistent effect on the subsequent course of allocation (Puffert, 2002). Path dependence thus explains why behavior once deemed rational can persist over time even if it is not deemed rational anymore. For example, the Stephenson gauge of 1435mm is, after almost 200 years, still in use today in most western railway systems although other gauges might be equally or even more efficient.

Path dependence can be explained using the model of Geels (2004) in Figure 3.1. The cumulative causality between actors and institutions is deemed one of the defining mechanisms behind the formation of 'routines' (Nelson and Winter, 1982) and the subsequent technological trajectory of a system (Dosi, 1982). A technological trajectory is defined as "the pattern of normal problem solving on the ground of a technological paradigm" where a technological paradigm is understood as a "prescription on the directions of technical change to pursue and those to neglect" (Dosi, 1982: 152). These prescriptions, the basic component of a routine, thus result in a certain direction the system tends to grow. Besides path dependence through routines, more formal institutional and economic factors play a role as well. First of all human actors, organizations and social groups learn from sociotechnical systems and build up cognitive capital that represents sunk costs. For firms it is often hard to develop competence-destroying breakthroughs (Tushman & Anderson, 1986; Christensen, 1997). As learning is a cumulative process, the longer this process takes place, the harder it becomes to accept innovations that make this knowledge obsolete. Furthermore, normative and formal rules make the system stabile (Geels, 2004). However, legally binding contracts (Walker, 2000) and technical standards (Geels, 2004) make a system also prone to inertia. On top of that, railway systems are highly capital intensive (Ksoll, 2004) and this causes path dependence in two ways. Firstly, capital-intensive industries are often characterized by a natural monopoly (Ordover et al., 1994) since high costs must be made upfront and this poses an entry barrier for potential competitors. An example would be the infeasibility of new entrants investing in a parallel railway network to compete with the Dutch railways. Incumbents then often have the incentive not to change the system, changes that might endanger their power (Geels, 2004). Second, high capital intensity makes sociotechnical systems rigid because of sunk costs. Therefore the path the system will go is determined by historical investments that need to be recovered. Transformations in infrastructures are therefore highly incremental (Markard, 2006) and "train systems and breakthrough innovations would appear to be a contradiction in terms" (Bontekoning & Priemus, 2004: 342)

4.1.3 Open-assembled systems

Whereas most technological artifacts are only to a limited extent subject to public scrutiny, networked infrastructures are a vital backbone to nations' economies (Herder et al., 2008). More than solely profit, these systems need to represent a broad set of public values such as sustainability, safety, accessibility and mobility (Thissen and Herder, 2003). Hence, these systems can be called open-assembled systems: systems where organizational, political and power dynamics to a large extent determine technological progress (Tushman and Rosenkopf, 1992). Explaining innovation processes for these types of systems thus needs to incorporate the notion that not solely technological rationality explains the dynamics but that dynamics located elsewhere have a pervasive effect. This creates a delineation problem, a recurring problem for system-level research (Ackoff, 1971; Tushman and Rosenkopf, 1992). For instance, governments regulating transport networks might be considered part of the overall system if one acknowledges its impact on the evolution of these networks. However, for someone interested in the development of a specific technology within the system, one might perceive government influences as external.

4.1.4 Multi-Level systems

That railway systems are full of human actors, operators, designers and managers gives rise to another feature of these systems. Being goal-seeking and intelligent, they are able to adapt to changing circumstance (Wooldridge and Jennings, 1995). Given complex systems comprise many of these adaptive agents, the combined interactions give rise to emergent phenomena, phenomena which cannot be related one-to-one to lower-level behaviors. The field of complex adaptive systems (CAS) therefore provides a good theoretical framework to study railway systems since it takes into account this complexity of systems and the adaptivity of its constituent elements (Holland, 1992). According to Holland (2006), CAS share four distinctive features: first, CAS consist of many elements (called agents) that interact simultaneously by sending and receiving signals. Second, these agents act conditionally on the signals they receive. Third, agents use modules of decision rules so they can react to new situation be recombining these sets of decision rules. Fourth, agents adapt to improve performance according to a preset 'credit assignment'. Agents assign credit to system outcomes and adapt so as to maximize this credit (Holland and Miller, 1991). This adaptation on an agent level occurs locally therefore a CAS is ever evolving. The resulting perpetual novelty leads to CAS never reach a global optimum or attractor (Holland, 1992). Immune systems, economies and markets can be classified as complex adaptive systems (Holland, 2006) but also railway systems can be seen as such (Herder et al., 2008). For studying complex adaptive systems, three levels can be distinguished (Bekebrede and Meijer, 2009): the level of the agents, the level of their interactions and the level of the emergent behavior. To some extent this relates to the strategic, tactical and operational decision making distinction often used in decision-making research as we show in Table 4.1:

Table 4.1. Decision-making and complex adaptive systems

Level of decision making	Level of Complex Adaptive System
Strategic	Observing emergent properties of system and determination of desirable attractors
Tactical	Design of interactions between actors and between actors and technologies to steer system towards attractor
Operational	Micro-level actor decisions

The complex adaptive systems perspective has been mainly used to study self-organization in autonomous collections of agents. In the infrastructure domain the framework is hence valuable to study so-called inverse infrastructures (Egyedi and Mehos, 2012) such as Wikipedia where infrastructures grow bottom-up. Still, for more traditional infrastructure systems such as railway systems, where to some extent top-down development is happening the framework points to two important features:

1. There is no one-to-one relation between emergent phenomena and lower-level behavior. It is therefore a complex task to design lower-level behavior to shift the system from one dynamic equilibrium to another.
2. Even if possible, possibility of top-down design is limited since agents at the operational level are adaptive. They are able to circumvent or make undone changes from higher levels.
3. Given this, top-down strategic decisions create many unforeseen consequences.

Concluding, we can say that railway systems are all of the above. They are complex in that both the structural and spatial configuration relate non-linearly to functional performance. They are sociotechnical in that human operators, designers and decision makers, organizations and institutions work together with technical artifacts in order for the whole system to be able to function properly. They are open and pervasive to political, environmental and contextual influences. And they are systems with adaptive agents that in their conjunction lead to emergent system properties. Therefore, any valid innovation framework able to provide any guidance on the analysis of temporal processes revolving around such systems should acknowledge these features.

4.2 Innovation

For any scientific undertaking, the knowledge starts quite declarative. Definitions and taxonomies encapsulate the object of study and leave out those phenomena that are not part of it. To define innovation, we first bound ourselves to the study of the process, rather then the product. Hence we adhere to Van de Ven's definition of innovation as: "*the development and implementation of new ideas by people who over time engage in transactions with others in an institutional context*" (Van de Ven, 1986: 604). This definition acknowledges two things: firstly, innovations happen over time and therefore are temporal phenomena per se. Secondly, innovation is both a cognitive undertaking where innovation actors develop ideas as well as a social undertaking where the aforementioned cognitive work has to be distributed over multiple actors, giving rise to transactions and institutions. We also add here that innovation is a multi-level process. As Marquis (1969) states:

"Keep in mind that innovation is not really a single action, but a total process of related subprocesses. It is not just the conception of a new idea, nor the invention of a new device, nor the development of a new market. The process is all of these things acting in an integrated way toward a common objective- which is technological change" (pp. 28-37).

According to Van de Ven's definition (1986), four factors are important in describing innovations processes: ideas, people, transactions and context. Innovation research thus far has only slightly taken on a holistic perspective, studying the interplay between these factors over time. Either a technological (idea) primacy (see for instance Abernathy and Utterback, 1978; Sahal, 1981; Tushman and Anderson, 1986, Kauffman, 1993; Frenken, 2000) or an organizational (people and transactions) primacy (Williamson, 1975; Pisano, 1990; Van de Ven et al., 1999) has been applied in studying technological change. However, dynamics in both realms are highly interrelated (Burns and Stalker, 1961; Leavitt, 1965; Hage and Aiken, 1969; Galbraith, 1977; Leonard-Barton, 1988; Henderson and Clark, 1990): when the innovation artifact changes, this leads to changes in the arena of involved human actors, and when the arena of involved human actors changes, this has consequences for changes in the innovation artifact. This partially explains the non-linearity of innovation processes (Kline, 1985; Cheng and Van de Ven, 1996, Anderson and Joglekar, 2012).

Innovation started as field of study since the works of Schumpeter (1934; 1950) showed its importance for economic development. According to him, it was the innovative capacities of firms, especially larger firms, and the resulting creative destruction of obsolete technologies and firms that caused economies to become ever more efficient in the production of goods and services. Since then broadly three communities have paid attention to the phenomenon of innovation: the economic, the technology management and the organizational sociology communities (Gopalakrishnan and Damanpour, 1997), each with a different focus on a particular stage of the innovation process, a different innovation type and different units of analysis. For this reason, until recently research on innovation has been highly fragmented (Tidd, 1997). As we will show, two theoretical frameworks have bridged these three communities and will serve as the lenses by which the further empirical part of this thesis will look at innovation processes in the railway sector. These frameworks are the Technological Innovation Systems framework (Carlsson and Stankiewicz, 1991) and the Multi-Level Perspective (Geels, 2002). Specifically these frameworks capture the multi-level dynamics found in innovation processes and allow for the structuring of process studies looking at otherwise chaotic and messy processes that characterize innovation (Cheng and Van de Ven, 1996; Anderson and Joglekar, 2012).

4.2.1 Innovation as a process

To show where current theoretical thinking on innovation stands and explain how it got there, we need to tell a story of two separate narratives. The first narrative is that of the economics sciences seeing the importance of innovation and of technology management trying to find ways by which to improve innovation processes in firms and sectors. The second narrative is one of new theoretical notions coming from the system sciences and inductive empirical work done on innovation processes inside firms. The core of the stances both narratives take is that between linearity and non-linearity. Whereas linearity functioned as a practical assumption for economic theorizing and a necessity for causal research, non-linearity seemed pervasive throughout in-depth case studies and a key characteristic of more general feedback systems.

Firstly, economical and technology management thinking originated from the notion that innovation was key for organizations to survive and for economies to prosper (Schumpeter, 1950). With a highly prescriptive flavor, this stream of research took on a highly linear perspective on innovation. Although, from a dynamic perspective, innovation was seen as crucial to both firms and economies, research zooming in on this phenomenon was portraying innovation as a linear process from invention to implementation.

Much of the earlier work on innovation was interested in the link between scientific research and applied inventions and studied through what stages this unidirectional link (Godin, 2006) worked. This link was perceived as orderly and linear, in that knowledge attained through scientific research would result in inventions through a set of orderly stages with little to no iterations (Bush, 1945). From the 60s onwards, the involvement of economists and management scholars in this debate led to the introduction of the concept of 'diffusion' as an important part of the innovation process. This led to a richer picture of how knowledge and new products were related, albeit still assumed in a linear fashion. Typical process descriptions of that time distinguished problem recognition phases, idea formulation phases, problem-solving phases, implementation phases and diffusion phases (Myers and Marquis, 1969; Utterback, 1974). The stage-gate-model (Cooper, 1990; 1994), popular still today in innovation management literature, falls in this category of linear models as well, albeit that iterations and parallel activities within stages are possible (Cooper, 2008). According to Cooper (1994), innovation processes move from idea generation stages through investigation, development and testing phases, to final market launch. The core of his prescriptive model deals with the management of uncertainty and investments. As an innovation moves through these phases, uncertainty tends to decrease and the need for investments tends to increase. Therefore, as the model shows, gates are put in place to determine whether or not an innovation deserves continued investment by the organization.

Innovation processes in organizations are likewise designed where steering committees govern these processes through determining whether or not an innovation project deserves further managerial effort and financial investments.

Linearity still pervades much of the innovation management literature, not so much for ontological reasons but more for methodological reasons. Large part of the academic work on innovation focuses on linking innovation variables to innovation performance in order to better understand, predict and improve innovation processes (Damanpour, 1991; Anderson et al., 2004). For this purpose, a highly static and linear perspective is taken for this allows the use of methods such as regression and variance analysis to uncover how variables correlate.

4.2.2 A systems' perspective on innovation

From the fields of physics and theoretical biology, the notion appeared that many features of a system couldn't be explained by looking solely to its constituent parts. The study of systems as a separate field of inquiry started with the seminal work of Von Bertalanffy, an Austrian biologist unsatisfied with the scientific paradigm of reductionism and Cartesian analytical thinking. In his thinking, he went back to the neo-platonist paradigm of the 16th century (Weckowicz Thaddus, 1989). According to Von Bertalanffy (1968), scientific progress could not be guaranteed by the analysis of causality in artificially created isolations. He argued that the behavior of the 'whole' could not be summed up by its isolated parts. His propositions mainly stem from his biologist background where at the time a debate arose about mechanistic and vitalistic views on life (Drack, 2009). Organisms seen as systems were either seen as a sum of inorganic sub elements related to each other deterministically and according to physical laws or seen as an atomic system for which biology needed other methods and laws. He asserted that indeed organisms merely exist out of lifeless matter but that the specific organization of this lifeless matter was the basic trait of 'living' (Von Bertalanffy, 1934 in: Drack, 2009). Following this notion, studies on systems started to concern the control of systems in Cybernetics (Wiener, 1949; Ashby, 1956), the behavior of causal systems over time in System Dynamics (Forrester, 1958), self-replication of systems in Cellular Automata (Von Neumann and Burks, 1966) and self-organization of system elements in Complex Adaptive Systems (Holland, 1992).

Most of these connotations of systems stress cumulative causation, sensitivity to initial conditions, non-linearity and path dependence as defining characteristics of complex systems. A systemic perspective would therefore cause any linear approach to studying innovation to be invalid since it neglects the potential deterministic chaos and path dependence that systems tend to have. Such features of systems are rarely describable using linear models. Problems with the linearity of many innovation process models were apparent, even at the time these models increasingly became popular in the 60s. Schmookler (1962) introduced demand-side aspects as important in innovation, thereby revoking the uni-directionality of the standard models. Rather then being the result of solely a technology-push from scientific research to diffusion, demand played a key role in directing scientific research and in directing innovative activities of firms. In addition, Price and Bass (1969) pointed to the challenges of organizations in bringing about radical innovations as they assumed organizations need internal changes, which cannot be anticipated beforehand.

The increasing amount of critiques of linear descriptions of innovation processes was also partly caused by management scholars focusing on how decision makers actually make decisions and what actually went on in organizations. Key notions in this time are Herbert Simon's satisficing decision maker (Simon, 1956), Cyert and March (1963) behavioral theory of the firm, Cohen, March and Olson's (1972) garbage-can decision making model and Lindblom's incrementalism (1979). Although not directly focusing on innovative efforts of organizations, the stream of new theoretical concepts revolving around the notion of fallible decision makers in political and messy organizational processes pointed towards the need to more closely link theory development with empirical research. Van de Ven's work with his Minnesota Innovation Research Project (MIRP) can be seen as one of the first attempts at inductively studying innovation processes (Van de Ven et al., 1999), staying close to empirical findings without forcing any theoretical framework beforehand. The less theoretically informed and more inductive approach resulted in a plethora of findings that contradicted simplistic linear models. A few stylized facts on innovation processes in organizations are that they are technologically complex processes that are social as well and include have a political and institutional dimensions.

4.2.3 Innovation as a technologically complex process

A complex system, as Simon (1969: 468) defines it, is "one made up of a large number of parts that interact in a nonsimple way". Because of this feature, system performance is likely to be nonlinearly related to changes in one or more elements of this system (Saviotti and Metcalfe, 1984; Davies and Hobday, 2005; Ethiraj and Levinthal, 2004).
These interdependences create the need to constantly take into account the part-whole relationships (Van de Ven, 1986). Innovation processes are therefore characterized by emergent irrationality, even if innovations on the part-level are deemed rational. The relation between design decisions and technology performance are highly uncertain.

For the individual decision maker this poses a problem. Decision making in these circumstance might be described as complex problem solving (Funke, 1991) given that these systems are complex, often only symptoms can be seen, and show many of their features in a dynamic context. The field of complex problem solving emerged out of psychological decision making research in the 1970s and represented a shift from simple puzzles to people solving complex problems (Fischer et al., 2012). Also this problem solving is a longitudinal phenomenon itself as for more strategic decisions, understanding deepens over time (Mintzberg et al., 1976). This is because, for radical decisions such as strategic and innovation decisions, decision makers can rely to a lesser extent on historical trends and routines (Mintzberg, 1973). They therefore face more uncertainty than managers (Hambrick and Crozier, 1985). Hence innovation management can be seen as the management, and reduction, of uncertainties over time (Kline and Rosenberg, 1986). Agreeing with Van de Ven (1999): "*learning is considered to be a key aspect of the process, where 'learning by discovery' is understood as "an expanding and diverging process", and learning by testing as "a narrowing and converging process" (Van de Ven et al., 1999: 203).*

The temporal phenomenon of innovation causes consequentiality of decisions (Simon, 1960). Decisions namely impact the range of feasible decisions later on. Agreeing with Brehmer (1992) decision-making in such a dynamic context is more process control than one-off decisions. Next to the consequentiality of decisions, problems themselves have a certain eigendynamic (Funke, 2003) meaning that problems change independently from the actions of the problem solver. Hence the individual problem solver needs knowledge about the structure of the system as well as the dynamics (Funke, 2001). Usual problems for process control than also hold for this dynamic decision making, such as poor feedback quality and feedback delay.

Given this uncertainty individual decision makers use other coping strategies besides learning. People employ heuristics to shortcut the search for a solution (Newell and Simon, 1972). Heuristics are procedures that help in finding satisficing solutions in lesser time than the time needed to find the most optimal one. Heuristics might come in the form of routines (Nelson and Winter, 1982) that have evolved over time, outlived other less valuable routines, and hence have embedded in them structural information about the to be solved problem. Heuristics are for instance to recognize the problem as a typical problem and use an old solution on this somewhat similar problem (Klein, 1989) and to use analogies for reasoning (Steinbruner, 1974). Heuristics are however fallible. Where this is the case, biases tend to appear. In the 1970s and 1980s much research has been done on these biases (see Hogarth and Reder, 1987 for an overview). Noteworthy biases are the 'prior hypothesis bias' (Levine, 1971), 'representative bias' (Tversky and Kahneman, 1975), escalating commitment (Staw, 1981), the rigid use of assumptions (Mason and Mitroff, 1981), the illusion of control (Langer, 1975) and the preferential bias for complete information when evaluating alternatives (Yates et al., 1978).

Biases happen however not always at the same time and for the same reasons. This is because decision makers can apply different strategies, and enforce different heuristics for different problems. The naturalistic decision making fields have highly contributed to the knowledge of these strategies. Based on analyzing 102 empirical cases Lipshitz and Strauss (1997) found five general strategies humans apply to cope with uncertainty. The RAWFS acronym stands for Reduction, Assumption-based Reasoning, Weighing, Forestalling and Suppressing and is depicted in Table 4.2.

Table 4.2. Five different uncertainty coping mechanisms

Strategy	Explanation
Reduction	Seeking additional information, seeking advice or delaying
Assumption-based reasoning	Using assumptions to fill in gaps, using analogies
Weighing	Rating different alternatives on their pros and cons
Forestalling	Building additional courses of action, creating slack and redundancies
Suppressing	Neglecting uncertainty; taking risks

As we have noted, decision-making in a context of innovation is highly dynamic and resembles much a sort of process control. A key factor here is then that innovation processes tend to be best described, from a technical perspective, as a constant interaction between technology and innovation actors (Thomke, 1998; Orlikowsky, 2002) since control processes and the control process are highly related (Brehmer, 1992). This interaction, in the form of design-test-analyze-design cycles, leads to the dynamics in the technical artifact and the knowledge of those who design it to be highly interdependent creating path dependence in innovation processes. Furthermore, knowledge by which complexity might be reduced is in itself highly path dependent (Rosenberg, 1976; Pavitt, 1984; Cohen and Levinthal, 1990), meaning that firms are unlikely to evaluate all possible innovations but rather focus on those that lie within the cognitive grasp of the involved actors.

4.2.4 Innovation as a social process

The complexity of innovation means that no single designer or decision maker is able to be solely responsible. In fact, rarely a single firm can said to be responsible for a specific innovation (Brusoni and Prencipe, 2001). Innovating complex systems creates the need to collaborate and coordinate with multiple actors for they bring to the game the needed resources, responsibilities and knowledge (Kanter, 1988; Stevenson and Jarillo, 1990; Brusoni and Prencipe, 2001). Consequently, innovation is not solely a technical design task but additionally one where the mobilization of resources in a network becomes key (Hounshell and Smith, 1988; Hoholm, 2011). Therefore innovation processes are highly impacted by the interactions between multiple designers, decision makers and other stakeholders (Fagerberg, 2005). In addition, innovation is not solely a process impacted by the suppliers of the technology. In understanding the dynamics of innovation processes, the role of users becomes highly important (Von Hippel, 1986), where user preferences and technology dynamics are interlinked (Pinch and Bijker, 1984).

4.2.5 Innovation as a political process

In retrospect innovations seem self-evident. However, similar to implementing new policies, many political dynamics characterized the innovation process (Hardy, 1994; Brown and Eisenhardt, 1995; Latour, 1996). Whereas March and Simon (1958) portrayed innovation mostly as a search process, Cyert and March (1963) behavioral theory of the firm added the theory of choice. Firms, albeit often assumed by economic scholars to be politics-free, work as governments as well (see for instance Dalton, 1959; Crozier, 2009) where advocates for different solutions have to bargain for their choice to be taken instead of to collectively solve problems (Pondy, 1967). As such, latent conflict is an everyday reality for firms, just as it is for governments (Pettigrew, 1992). Inside the innovation process many controversies arise about where the innovation should be heading (Van de Ven et al., 1999; Beunza and Stark, 2004; Howard-Grenville and Carlisle, 2006). In addition multiple innovations compete for resources simultaneously. Innovation is therefore also a process of network building around ideas and worldviews (Latour, 1987) in opposition to other ideas and worldviews.

4.2.6 Innovation as an institutional process

Technology and the institutions around it are highly interrelated (Williamson, 1973; Teece, 1996; Geels, 2002). Given that innovation is a process undertaken by fallible decision makers with differing goals, incentives and frames of references, transaction costs arise when these decision makers need to cooperate and share resources (Burns and Stalker, 1961; Lawrence and Lorsch, 1967). Hence, innovation is not solely a technical process, but also a process where the management of transactions becomes key (Van de Ven, 1986). For this reason, institutions appear, either consciously designed or emergent, to reduce the social uncertainty in transacting with fallible and opportunistic actors. On the other hand, institutions also impact what technological opportunities are perceived and taken (Nelson and Winter, 1982; Dosi, 1982). In these collaborative settings, institutions can be defined as the rules of the game (Nelson and Nelson, 2002). These rules can be norms, laws, contracts, cultures, etc. In an innovation context, routines have played a large role as institutions creating technological trajectories (Nelson and Winter, 1982). Routines are recipes of a programmatic nature, which guide individual behavior and said to largely determine where actors involved in technology development look for opportunities for further improvement.

4.2.7 Overall

In general, all these accounts of innovation processes from different perspectives have led to a picture of chaos, contingency, and controversy. Innovation is therefore an uncertain process (Van de Ven and Polley, 1992). Innovation consists of multiple processes carried out by many actors through complex social interactions that change over time (Hoholm and Araujo, 2011). Technology and organizations and users constantly co-evolve (Hage, 1980; Leonard-Barton, 1988), meaning that a sole focus on technology would diminish explanative power of innovation research highly as human actors and institutions should be endogenous factors, even those outside the direct scope of the single firm (Chesbrough, 2003). Three key notions therefore play a large role in subsequent theorizing on innovation processes:

1. Statically, innovations are systems of interdependent technical, social and institutional systems
2. Statically, innovations are multi-level phenomena. It is both about micro-level actor behavior and macro-level emergent patterns.

3. Dynamically, the dynamics in each subsystem and level constantly are influenced by and influence other subsystems and levels. This creates feedback, chaos and cumulative causation.

Linear models on innovation processes do little justice to these notions of systemicity and dynamics. For both scientific rigor as well as practical relevance, such a linear focus is worrisome. Firstly, chaotic, multi-level and multi-dimensional and path dependent processes are hard to capture using simple linear models. Secondly, actors involved in innovation processes are rarely helped with knowledge about how innovation performance is increased by increasing for instance the multi-disciplinarity of the design teams. Such knowledge provides little guidance as to how to achieve this, when and when not, and with whom.

4.3 Current models on innovation

Given, however, that most of the innovation research still focuses solely on relating independent variables to dependent variables such as for instance innovation speed and innovation quality (Anderson et al., 2004), a structured understanding of innovative behavior in firms is still underdeveloped (Wolfe, 1994; Salaman and Storey, 2002) and no single general theory on innovation is to be expected. Adhering to the notion of complex systems not being describable by one formalism (Mikulecky, 2001), there is also no sense in trying to find one. Rather, the analysis of innovation in these systems should be based on looking at it through a plethora of perspectives and with a plethora of different methodologies. We present an overview of common models used to describe and understand innovation in system-like technologies.

4.3.1 Large Technical Systems

The evolution of networked infrastructures is a central topic in literature on large technical systems (LTS, Hughes, 1987). Before the 1980s, the study on these large systems has mainly been undertaken from an economic historical perspective, focusing on its growth, its economic implications and its management, often with the technology itself as given (Joerges, 1988). Starting from the works of Hughes (1983, 1987) the field of Large Technical Systems emerged that sought to combine the works on large infrastructure systems with the studies on history of technology, a field that mainly focused on individual inventors and singular technical implements (Joerges, 1988). Unique to LTSs is that these systems are large in the sense that they are both global in scope and structure (Gökalp, 1992), in scope because most people within a country are affected by them, in structure because many factors contribute to the working of these systems. Simpler put, LTSs are infrastructure networks that stretch geographical areas (Geels, 2007). Studies on these LTSs have mainly focused on describing how these technical systems evolve over time and what factors influence their growth. Basically, the evolution of LTSs can be divided into an initial phase, an acceleration phase, a stabilization phase and a decline phase (Gökalp, 1992). Hughes' framework of LTS revolves around two main concepts (Hughes, 1987) and places the system builder on the foreground. First of all, growth in these systems happens around ' reverse salients', anomalies that result from an unbalanced evolution of the system: progress in one end may hamper progress elsewhere in the system. Large systems grow to such an extent that they create momentum: the invested interests of many people and organizations in the system and the fixed assets and sunk costs give it mass, directions or goals and display a rate of growth. For our purposes however, this framework provides only little value for two reasons: firstly, the framework focuses on overall system evolution and neglects more revolutionary change in parts of the system. Secondly, the framework awards the system-builder a central spot, neglecting more bottom-up processes (Egyedi and Mehos, 2012).

Three models appear to do justice to at least one of the three central notions on innovation processes. Within the more technical domain, NK-models (Kauffman, 1993) portray complex technologies as bundles of elements mapped into fitness functions and the resulting topology of fitness values. These models are a-social in that human agency is reduced to simple algorithms by which agents search a fitness landscape for better configurations of a complex system. Agency and institutions play a larger role in Technological Innovation Systems (Carlsson and Stankiewicz, 1991) and the Multi-Level Perspective (Rip and Kemp, 1998; Geels, 2002). Whereas the first helps in describing the dynamics of those closely involved in the process, the latter helps in describing more general transformative patterns in system innovations. However, both neglect to a certain extent the technology part and its pervasive effect on innovation processes. We therefore feel that the three perspectives perfectly counter each other's weaknesses.

4.3.2 NK

Kauffman (1993) first started to use complexity as a way to describe biological evolution, advancing the work on fitness landscapes started in evolutionary biology by Wright (1932). His NK-model portrays complex systems as having N elements and K interrelations with each other, together determining the fitness of the system. Fitness, in his model, is a one-dimensional construct related to how well a system is able to perform in its environment. N elements can have multiple states (A) and the combinatorial parameter space (A^N) is therefore seen as the design space: it is the totality of all possible configurations of system elements. For instance, a system having three binary (0 or 1) elements has potentially 2^3=8 different configurations. The sole focus on elements and interrelations allows for an abstract representation of any complex system, being biological systems, technological systems, or social systems.

In this model, evolution of any system is represented as a recombination of existing elements through processes of variation, selection and retention. For biological systems these are ultimately the four nucleotides that through their specific grouping create genes and ultimately different traits that either hinder or support survival and reproduction.

Originally intended to model biological systems, the model has been used to describe technological systems as well (e.g. Kauffman and Macready, 1995; Levinthal, 1998; Fleming and Sorenson, 2001; Frenken, 2000; 2006). Technological systems are, just as biological systems, a collection of elements that together are in connection to an environment. Innovation can then be seen as the mere reshuffling of already existing elements (Fleming and Sorenson, 2001; Frenken, 2006). It therefore provides valuable insight into change processes in the technological domain of networked infrastructures. Most often, these change processes involve the recombining different already existing elements. For instance, overhauling the track layout together with different timetables, operator roles and rules and new procedures for delay mitigation can be classified as a radical departure from the norm yet do not involve anything totally new.

As an example, consider a car that consists of only an engine, a braking system and tires. All three of them can have two states: electrical engines or combustion engines; basic braking systems or advanced braking systems; basic tires or advanced tires. In this hypothetical example, 8 different cars can be made. The complexity here involves the interrelations between these elements in determining how 'fit' the car is. Choosing a higher-powered combustion engine will influence both directly the fitness as well as indirectly through changing the influence of other elements. The power of the engine has a direct effect on fitness (the car has a higher speed) but also an indirect effect (the tires, at higher speeds, perform less well).

These interrelations between elements in determining the overall fitness of the system are termed epistasis. Theoretically epistasis (K) can range between 0 and N. When a system has an epistasis of K=0 all elements independently from each other impact fitness. Changing one of the elements of the system to increase the fitness of the system will not cause any second order effects on the impact of other elements. For maximum complexity, when K reaches value N, all elements are related to each other and optimizing for fitness then needs to incorporate all moderating effects on the influence of one element on overall fitness. To incorporate the multi-dimensionality of the concept of 'fitness' Altenberg (1995) put forward a generalized version of the NK-model. His model allows fitness to be defined by a range of different fitness properties, rather than a single value. In our previous example, the fitness of a car is then broken down to traits such as speed, energy efficiency and cost.

The generalized NK-model then conceptualizes fitness and system configurations as the phenotype and genotype of a system. Phenotypes are the discernible traits of a system that in their conjunction make up the fitness of the system. Genotypes are the underlying structural coupling of system elements. Instead of its epistasis, the complexity of these systems is found in two defining characteristics of these systems: their pleiotropic and polygenic properties. Pleiotropic properties mean the extent to which a system elements' contribution to a phenotype depends on other elements of the system. For instance, speed of running of organisms is determined by the specific configuration of many different genes together. Polygenic properties mean the amount of different phenotypes that a single element has an influence on. Thus in portraying these systems and its complexity characteristics, genotype-phenotype maps are an instructive way to visualize the systems' internal architecture (see Figure 4.2)

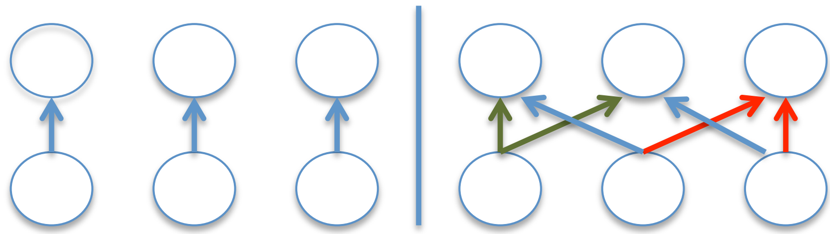


Figure 4.2: genotype-phenotype maps for a simple system (a) and a complex system (b); pleiotropy (red) and polygeny (green) are highlighted

Returning to the example of a system with N=3 elements and epistatis of K=2, a fitness landscape in the form of a cube would look like Figure 4.3:

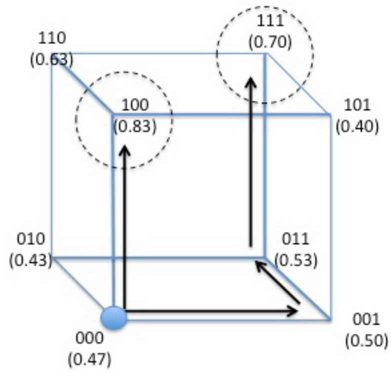


Figure 4.3. Fitness landscape of an artifact with N=3 (three elements) and K=2 interdependencies

Both configurations 100 and 111 are optima, points where single element changes will not cause any further improvement (Rivkin and Siggelkow, 2002). If 000 were the configuration with which the system started, myopic search would lead to either global optimum 100, or local optimum 111. This solely depends on the first decision as to change which of three elements. If 001 is firstly chosen, the system will grow towards state 111 after which no single element change will result in an improvement. 111 is optimal, but given a more optimal point elsewhere, this optimum is deemed local.

4.3.2.1 Fitness landscapes

Fitness landscapes then portray the topology of fitness values for all possible configurations of system elements. For systems with epistatis of 0, the topology would resemble Mount Fuji, where there is one global optimum where incremental innovation, changing one element at a time, would always result in reaching this optimal point. This is logical, since any one-element change does not result in other elements changing their contribution to the fitness of the system. For more complex systems, landscapes tend to be more rugged. The fact that individual elements' contribution to fitness is highly contingent upon other elements creates many local optimal points in the landscape. In Figure 4.4, we again portray three fitness landscapes for systems with increasing complexity:

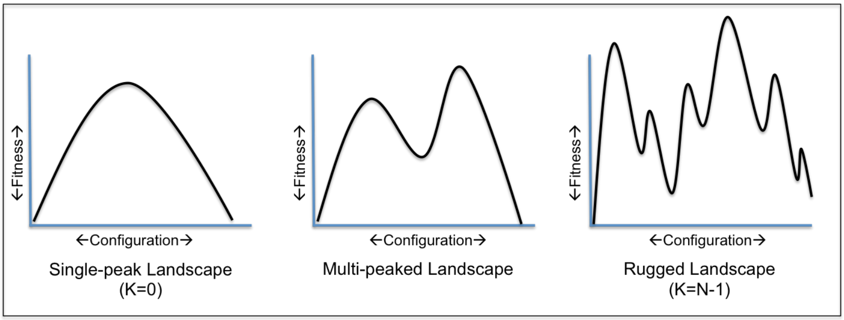


Figure 4.4 Fitness landscape for a simple product (left) and a complex product (right)

4.3.2.2 Search

In the social science domain, NK-models have been mostly used to study the effectiveness of different search strategies (Kauffman et al., 2000; Fleming and Sorenson, 2004; Knudsen and Levinthal, 2007). The basic premise is, agreeing with March and Simon's satisficing decision maker (1958), that innovators do not know the full topology of the fitness landscape. That is, resource constraints and cognitive limitations inhibit the innovator to exhaustively search the landscape for the most optimal point (Kauffman et al., 2000). The model is able to elegantly show how different product architectures and search strategies are related. The benefits of modular architectures are for instance a key topic within this stream of research (Ethiraj and Levinthal, 2004; Frenken and Mendritzki, 2012) as these types of architectures allow for myopic search to lead to fewer lock-ins in local optima. Other topics addressed are the role of complexity in preventing imitation by competitors (Ethiraj et al., 2008) and the design of organizations to manipulate the fitness landscape itself (Levinthal and Warglien, 1999).

In general, two parameters best describe the extent to which the landscape can be searched for more optimal configurations. These are breadth and depth (Katila and Aruja, 2002; Rosenkopf and Almeida, 2003). Breadth is the amount of system elements that can be changed simultaneously and hence can also be classified as the search distance (Frenken, 2006). The more elements that can be changed simultaneously, the further away from the current point in the fitness landscape one can look for improvement. In the car example, changing both the engine and the tires at the same time and testing the resulting fitness is seen as having higher breadth as just testing a car with different tires. Lower search distances mean that one can only search neighbouring configurations. These strategies are often called myopic or local search strategies (Frenken, 2006).

Given that search, and testing, is rarely a one-off activity (Thomke, 2001; Erat and Kavadias, 2008), depth involves the extent to which searcher can look at multiple points close to each other at the same time. Because rugged landscapes mean that neighboring configurations might have a completely different fitness value, higher search depths become important. Remember that ruggedness means that a slightly different configuration of the system does not mean a slightly different fitness value. For higher search breadths, or higher search distances, depth then involves the extent to which the search process is either a lucky shot on a distance place of the fitness landscape or a careful analysis of more wider areas of the landscape. To its extreme, it would involve totally uncovering the landscape which we earlier on termed exhaustive search.

For local search strategies, high search depth then means the extent to which a searcher changes all single elements, albeit one at a time, to choose the best incremental step. This search strategy is called 'greedy search' (Frenken, 2006). Low search depths, for local search, would involve the changing of single elements of systems after each other, without looking for the best single element change. This modular search strategy, if neglecting the probable non-modularity of the artifact, can create so-called problem solving oscillations (Mihm et al., 2003) where organizations keep on solving the same problems over and over. In Table 4.3 we distinguish four different search strategies according to their scores on depth and breadth.

Table 4.3: different search strategies for fitness landscape

	High breadth	Low breadth
High depth	Exhaustive search	Greedy search
Low depth	Lucky-shot	Modular search

Innovation is then a trial-and-error process where elements are changed and only those changes are made permanent that increase the fitness. In the literature on NK-models, most often search is deemed myopic. Because of this feature, complex systems tend to climb the nearest hill until no further improvements are possible. Then, a complex system is assumed to be locked-in into a local optimum. Representing the trade-off between conflicting constraints. For complex systems it then holds that the starting point of the development process highly determines which optimal point is reached. The basin of attraction of an optimum is the range of configurations that will eventually climb towards this optimum. In Figure 2, the simple system has one basin of attraction, covering all configurations. The complex system has four local optima, of which one is globally optimal and where the left two optima have the biggest basin of attractions. The borders between two basins therefore represent tipping-points: configurations that are slightly left from the border will not grow slightly different but hill-climb towards a totally different peak in the landscape.

4.3.2.3 Transitions

However, recall that innovations can be either autonomous or systemic and radical and incremental. Distinct from many of the applications of NK-models, radical and systemic innovations would involve many element changes in a coordinated fashion. Given that these innovations encompass multiple changes, the specific sequence of changes becomes non-trivial. The recombinatorial nature of innovation gives way to equifinality: many different paths will lead to the same result. Alkemade et al. (2009) provide more insight into this phenomenon by using NK-models to study transition processes in technological systems. Here transition paths, from one configuration to another configuration, are described as the sum of all transition steps. The study zooms in on the amount and type of flexibility that transition steps cause, given the fact that fitness estimations might change, certain steps on the way might become infeasible or that preferences shift. Respectively, the flexibilities are called: design flexibility, path flexibility and preference flexibility.

The NK-model provides an instructive way to elegantly describe the static and dynamic complexities of technological artifacts. The nomenclature provides us plenty of concepts to describe railway system development, such as: polygeny and pleiotropy, hill-climbing, tipping-points, search strategies, optima and different flexibilities in transition path planning. However, the model is highly abstract and neglects to a large extent the highly influential dynamics of organizations taking these steps. Hence a few notions appear:

1. Fitness is a multi-dimensional construct. Innovation is a multi-actor activity and fitness is a highly contested due to actors finding different dimensions more relevant.
2. The valleys represent tipping points, after passing such a point autonomous innovations will lead to hill-climb this new optimum. However, before reaching such a tipping-point, innovations tend to decrease the fitness rather then increase. Therefore vision-setting, strategizing and institutional arrangements become highly important.

Where NK-models shed little to no light on these phenomena, other more 'social' frameworks focus specifically on the socio-institutional nature of innovation. These two are the Multi-Level Perspective (Geels, 2002) and the Technological Innovation Systems (Carlsson and Stankiewicz, 1991). Although the first focuses specifically on longer term societal transitions and the latter focuses on the diffusion of technologies as a whole (Markard and Truffer, 2008b) both provide valuable insight in systemic innovation processes on a smaller scale. As they have significant conceptual overlap (Weber and Rohracher, 2012) the combined use of MLP and TIS as a 'heuristic device' will improve the analysis.

4.3.3 Multi-Level Perspective

The Multi-Level Perspective (MLP; Rip and Kemp, 1998; Geels, 2002) is a theoretical framework that deals with transformative processes in sociotechnical systems. A sociotechnical system performs some sort of function and as such can be described as a configuration 'that works' (Rip & Kemp, 1998). However, given its path dependent development process, other configurations might work better. In terms of the NK-model, the framework would allow for the study of process by which sociotechnical systems move from one peak of the fitness landscape to another peak. As we have mentioned, such jumps are highly unlikely given the prominence of incremental innovations in large-scale systems. MLP frames these jumps as happening through the interplay between three distinct but related levels: the landscape, the regime and the niche and awards a pivotal role to niches. Given that the regime favors a certain set of incremental innovations, niches allow for the exploration of innovations defy the regime. The three levels differ on their structural and temporal scale (Raven et al., 2012).

Transition scholars use the MLP to study both historical and on-going transitions. Examples of historical transitions are the transition from sailing to steam ships (Geels, 2002), transitions in water supply (Geels, 2005a), electricity systems (Verbong and Geels, 2007), aviation technology (Geels, 2006) and the transition from horse-drawn carriages to automobiles (Geels, 2005b). Applications of this framework can also be found in the literature on on-going and future sustainability transitions (Nykvist and Whitmarsh, 2008; Van Bree et al., 2010; Geels, 2012). Furthermore highly related but more prescriptive frameworks such as Strategic Niche Management (Kemp et al., 1998) and Transition Management (Rotmans et al., 2001) deal with how to actively steer these transitions. As such all three perspectives can be seen as a revival of long range planning (Voss et al., 2009), albeit with a lower focus on substantial planning – of future solutions - but more on management of processes that could potentially lead to these future solutions. Given that all three perspectives are highly related and all three of them focus on niches as crucial for systemic innovations, we will use the theories interchangeably.

Broadening the cognitive focus of technological regime (Nelson and Winter, 1982), the MLP sees regimes as those structures that drive the individual behavior of elements of the sociotechnical system and hence exert a structuring force on the development of alternatives (Kemp et al., 1998; Smith et al. 2010). Rip and Kemp (1998: 338) define a technological regime as:

"the rule-set or grammar embedded in the complex of engineering practices, production process technologies, product characteristics, skills and procedures, ways of handling relevant artifacts and persons, ways of defining problems – all of them embedded in institutions and infrastructures"

The further move away from the technological regime notion from Nelson and Winter (1982), which had solely a cognitive focus, Geels (2002) introduced the term sociotechnical regime. Sociotechnical regimes consist of user practices, routines, shared beliefs, institutional arrangements and regulations amongst many others (Geels, 2011). Noteworthy here is to mention that sociotechnical systems as well as actor networks are not part of the regime. Rather, rules are embodies by technologies and routines are performed by actors and actor networks (Markard and Truffer, 2008). Other authors have however put forward that technology and actors should be considered part of the regime (Hoogma et al., 2002; Raven, 2007; Konrad et al., 2008). We adhere to this latter notion since many of the structuring forces of regimes are due to the technical characteristics of sociotechnical systems and due to the properties of actors and actor-networks.

The regime is emergent in that the institutional characteristics are the outcome of the collective action of many human actors over time. Changing regimes at will is therefore highly unfeasible (Kemp et al., 2001). In the reverse direction, institutions also structure collective action of human actors. The regime concept has therefore much conceptual overlap with the duality of structure notion of Giddens (1979). Human behavior is as much determined by as it determines emergent structures. Regimes in the MLP are thus both medium and the outcome of action (Ravens et al., 2012). Because of this mechanism regimes are not inert but are dynamically stable for without landscape pressures regimes tend to reproduce itself (Geels and Schot, 2007).

Landscapes can be seen as the external environment of regimes. Elements of a landscape are for instance broader cultural norms and values, economic conditions, demographic developments, oil prices, environmental problems or political dynamics. As such they are background variables (Kemp and Rotmans, 2005). That is, these variables significantly impact the dynamics of regimes (and niches) but are themselves not impacted by the regime. For a regime, the landscape provides the setting by which they have to optimize the sociotechnical system. Given that landscapes are highly structured and evolve over longer time spans than regimes, the dynamic stability of regimes cause these regimes to slowly adhere to these broader background variables. However, shifts in landscapes are possible. For instances rising oil prices might make a regime less effective, or rising concern over sustainability might demand a regime to adapt.

Niches are those places where the structuring forces of the regime are less prevalent. It is this concept that shows the evolutionary nature of the MLP. As regime elements have co-evolved over time, think of the interdependent evolution of technology, actors and institutions inside this regime, they are hostile towards all to radical change. Often incumbent regimes are not even able to truly value such a radical innovation. Transformative processes are therefore processes where both technology and use follow a fit-stretch pattern (Geels, 2005). Technology stretches need to be accompanied with stretches in user practices to fit each other. Geels (2005) provides a telling example about how automobiles were firstly very similar to horse-drawn carriages where the horse was replaced with a battery. Although the introduction of the battery formed a technological stretch, the stretch of use came later as users only later on saw the plethora of new opportunities caused by replacing horses with batteries.

Niches are then places of less structuration, green fields where radically different settings of a sociotechnical system can originate, explored and nurtured. It allows so-called change-inclined regime players to test radical innovations (Rotmans and Loorbach, 2009). Niches come in the form of Greenfield sites, pilot tests, niche markets, and laboratories. Their shielded properties allow for sociotechnical experiments, as the innovation within the niche does not have to answer immediately to the pressures of the regime. In other fields of study, niches might come in the form of economic protectionism for infant industries or medical incubators for premature infants.

Of crucial important is then that niches grow and finally coalesce with or replace the incumbent regime. This growth occurs through processes of learning, vision-setting and network building (Kemp et al., 1998; Hoogma et al., 2002). When these three processes are strong enough and influence each other, niches build up momentum (Geels, 2012). Momentum is seen as key to niche development as through this process they are finally able to resist selective pressures by the regime. Hence nurturing niches involves:

1. Coordinating learning activities to allow for the rise of a dominant design
2. Setting of visions that are inspiring and become accepted by both those in the niche as well as in the regime
3. Building of networks so as to create legitimacy through the involvement of powerful actors

The growth of niches is however not a phenomenon solely attributable to the dynamics internal to the niche. Although transitions are eventually sourced to these niches, they only occur in conjunction with processes happening at the level of the regime and the landscape (Geels and Schot, 2007). Landscape changes create pressures on existing regimes and in some cases this might lead to destabilization processes in the regime itself. This creates windows-of-opportunity for niches to grow and merge of invade the incumbent regime. An informative depiction of this multi-level process is Figure 4.5.

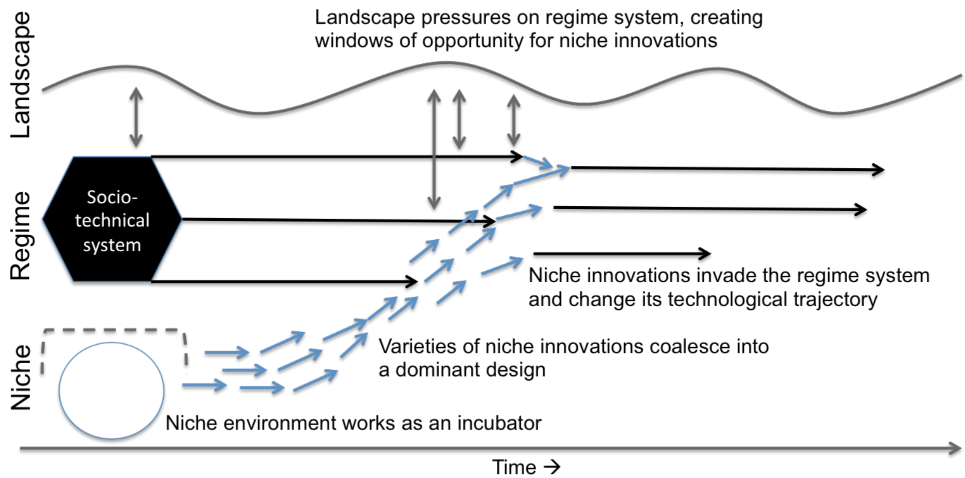


Figure 4.5 multiple levels during transitions (adapted from Geels and Schot, 2007: 401)

Depending on how this process is organized and by whom, different transitions might be distinguished (Berkhout et al., 2004). Regimes are able to some extent to use internal resources to adapt to outside landscape pressures. In these instances, they either reorient their trajectory through the emergent outcomes of uncoordinated efforts of regime players or through coordinated endogenous renewal. On the other hand, some forms of pressures are not answered to by current regimes. In these instances, change occurs through outside resources being put to use. When coordination is low, such as is the case for change through bottom-up growth of small high-tech firms, the transformation is emergent. When coordination is high, a transition is deemed purposive. Figure 4.6 shows this 2x2 typology.

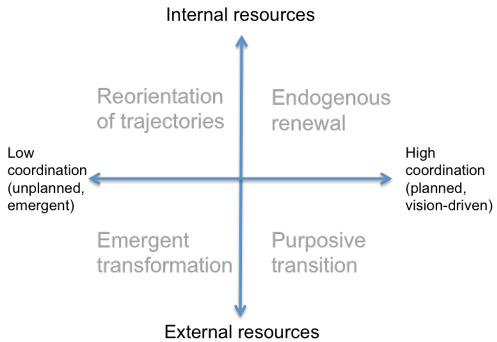


Figure 4.6 a typology of different transitions (Berkhout et al., 2004)

Agreeing with Geels and Schot (2007), the typology is not without its flaws. Especially the level of coordination is a doubtful parameter to distinguish transitions. As network building and vision setting are an integral part of niche growth, coordination is something that always emerges at later stages of transitions. Hence, transition processes contain both purposive and emergent elements. Geels and Schot (2007) designed a similar typology based on two parameters: the timing of interaction between the three levels and the nature of the interaction. Timing mainly involves the extent to which niche innovations are mature enough when landscape pressures arise. The nature of the interaction involves the extent to which both landscape pressures and niche innovations are in competition with the regime or in a symbiotic relationship. They subsequently distinguish five pathways:

1. Transformation: if a regime has to answer to disruptive change in the landscape while niches have not yet matured, they redirect innovative effort
2. De-alignment and re-alignment: if landscape changes are of an avalanche-type and no fully grown niche exists, the regime may break up. Different niches then compete for dominance
3. Technological substitution: if landscape changes are sudden and disruptive and niches are mature, these niches will replace the incumbent regime
4. Reconfiguration: when niches are in a symbiotic relationship with the regime, the regime may use the niche innovation to solve local problems. The introduction of this innovation as add-ons may trigger additional changes in the regime later on
5. Reproduction: when landscape changes are of a steady kind, the regime reproduces itself

Transitions are also not a story of one niche taking over the regime. An important feature of transitions is the interlocking of technologies (Geels, 2005). This process of interlocking creates additional dynamics to the innovation process and comes in four forms:

1. Complementary interlocking: two technologies in niches combine to strengthen each other
2. Hybrid interlocking: a technology in a niche combines with a technology in the regime
3. Sequential accumulation: a technology that invades the regime creates an additional window-of-opportunity for another technology
4. 'Borrowing': two technologies in niches might borrow elements from each other. They remain separate but now share more of the same elements.

Different from real evolution, transitions are to some extent intelligent processes. Firstly, search for new opportunities is a purposeful search activity (Frenken, 2006) rather than simple random variation. However, the model is often critiqued for neglecting agency (Smith et al., 2010; Genus and Coles, 2008). As a more global model, the MLP tends to focus more on patterns and less on micro-level behavior (Geels, 2011). Furthermore the model to some extent neglects power dynamics (Shove and Walker, 2007; Smith and Kern, 2009; Grin et al., 2011) and rational choice (Geels, 2010). This is troublesome, as we have shown how innovation process are both political process as well as process in which substantive design of artifacts take place and where both structural forces and agency play a role.

4.3.4 Technological Innovation Systems

That innovation is not solely a technical process but rather partly a social and institutional process is a notion also put forward by evolutionary economics scholars at the beginning of the 1980s (see Dosi, 1982; Nelson and Winter, 1982). An overarching framework for studying innovation on a national level that incorporated both technology and social reality was the National Innovation Systems framework (Nelson, 1992; Lundvall, 1998). This notion, where innovation systems are ensembles of co-evolving technologies, actor networks and institutions, later found fertile ground in the studying of economic sectors (Malerba, 2002) and for technological specific clusters of firms (Carlsson and Stankiewicz, 1991). The latter provides much value in studying innovation processes in railway systems, due to its delineation of the unit-of-analysis based on the technological artifact that is to be innovated. This latter approach, termed the Technological Innovation Systems, provides a framework to study the processes inside a 'niche' more in-depth. As such it significantly helps to add the micro-part to macro-part provided by the MLP.

According to Carlsson and Stankiewicz, 1991: 111) a technological innovation system is: "a network of agents interacting in a specific economic/industrial area under a particular institutional infrastructure or set of infrastructures and involved in the generation, diffusion, and utilization of technology".

The main goal of an innovation system is hence driving forth the innovation process and making sure the innovation is implemented (Edquist, 1997). However, the systems approach does not assume a collaborative effort (Bergek et al., 2008), but rather focuses on system dynamics that arise when both competing and cooperating actors in a changing network impact each other.

Studies taking on a TIS perspective have originally focuses on the structural analysis of innovation systems, looking at how structural features can lead to system failures (Carlsson and Jacobsson, 1997). System failures are events when an innovation system is not able to perform its functions such as the diffusion of a technology. Usually, failures are coupled with the subsequent dissolving of the innovation system. Klein Wolthuis et al., (2005) found four common structural reasons related to system failures. These are infrastructural failures, interaction failures, institutional failures and capability failures.

To better allow for studying why some innovation systems reached this goal and why some did not, many have delved deeper in the specific functions an innovation system should perform. In general, an innovation system is successful if it performs the following seven functions (Bergek et al., 2008; Hekkert et al., 2007; Hekkert and Negro, 2009)

1. Entrepreneurial activities

Innovations are not occurring autonomously. They need human actors willing to provide the needed energy to bring about changes. This energy makes sure that the potential of knowledge development, networks and markets is materialized.

2. Knowledge development

Of great importance are mechanisms of learning. Learning by doing and learning by search make sure that uncertainties about the functionalities of the innovation are taken away.

3. Knowledge diffusion through networks

Innovation is also a task of coordination, especially if many heterogenous actors are involved, such as public agencies, universities, research laboratories and private firms. The spread of information through networks is then key.

4. Guidance of search

Next to the coordinative effects of information diffusion, the innovation system should also be able to coordinate efforts through vision building and disciplining institutions. That is, all activities within the system should converge over time to provide better synergy between them.

5. Market formation

An innovation system should be able to create markets for itself to sustain the viability of the innovation. When existing technologies and existing markets are hostile towards these radical innovations, innovation systems should create niche markets or technological greenfields to keep the innovation away from the selective pressures of the regime system.

6. Resource mobilization

Next to niche markets and greenfields, innovations need to grow and therefore need financial and material resources. The system involved should be able to acquire these resources.

7. Creation of legitimacy

In the end the innovation system needs to replace or become part of an incumbent regime. Therefore it needs legitimacy to overcome resistance to change. Coalition forming is in this regard highly important.

In essence, the TIS framework is highly static. It provides a way to take snapshots of the performance of innovation systems according to how well such a system performs a set of 7 functions at a specific point in time. Because of this, the framework is highly valuable for cross-sectional comparisons, and accordingly the framework is often used to compare several innovation systems to explain why some were more successful than others. However, the cumulative causation between these seven functions allows innovation scholars also to better uncover the dynamics, albeit a solely functionalist one (Suurs & Hekkert, 2009; Bergek et al., 2008). Hekkert et al. (2007) explicitly calls for using the TIS framework in a more dynamic sense. In studying dynamics over time, they use allocation schemes to relate events on a micro-level to functions on innovation systems. However, two important features are missing. Firstly, as we have shown technology should be endogenous to the model since dynamics in technology and dynamics in the socio-institutional realms are highly related

(Markard and Truffer, 2008b; De Bruijn and Herder, 2009). Bergek et al. (2008) for instance describe how innovation systems can benefit from interacting with other innovation systems. This most certainly has an impact on the evolution of the technological artifact and subsequently the evolution of the innovation system around it. Secondly, the TIS framework is very much inward looking (Markard and Truffer, 2008b; Meelen and Farla, 2013), studying the specific dynamics of actor-networks over time inside the system. By doing so, it neglects the impact of regimes. Secondly, the framework lacks certain rigor in relating micro-level events to macro-level functions of innovation systems. That is, changes in the way the innovation artifact is made up, the way actors are working together or the way institutions are shaped over time are not related to functions, e.g. market formation, through a rigorous translation scheme.

4.4 MLP, TIS, and NK combined

All three models focus on different aspects of innovation processes. NK-models, disregarding the social and institutional realities, focus mainly on the technical complexities of innovation. On the other hand, more social and institutional theories such as the MLP and TIS, specifically allow for the study of the behavior of actors and institutions. The approach both frameworks apply is however different. MLP is more structuralist and dynamic in that it frames innovations as long term patterns evoked by micro-level behaviors. TIS zooms in on the micro-level and portrays the coupled dynamic between actors and institutions and their relation to a set of seven vital functions.

Since both MLP and TIS to some extent neglect technological dynamics, the addition of the NK-model points to a few additional notions:

1. The NK-model shows within innovation processes there is equifinality
2. Also it shows that the sequence of manipulation of elements matters
3. This means that during an innovation process, the innovation itself changes
4. It also means that because of different paths towards final implementation there are also different 'ups and downs' in system fitness through time.
5. These different ups and downs are perceived differently by different actors and over time their perception might change due to new evidence or changing incentives

In Table 4.4 we show each framework's contribution to the analysis of innovation processes as well as its flaws:

Table 4.4. How different frameworks complement each other

	Pro	Cons
NK (Kauffman, 1993)	Allows for studying complex technological systems and the dynamics related to the internal architecture of the system	Disregards social, political and institutional realities
MLP (Geels, 2002)	Allows for the studying of radical innovation processes over time over multiple levels of structuration, takes into account the battle between the new and the incumbent; acknowledges trade-off between exploration and exploitation (Levinthal, 1990)	Neglects micro-actor behavior, especially related to power and rational choice. Serves as a meta-theory, hence needs additional micro theories of design and decision making in networks. Highly structuralist in its approach
TIS (Carlsson and Stankiewicz, 1991)	Bridges micro-actor behavior and emergent system functions. Functions are described in detail and validated.	Static approach, neglects role of technological change in determining overall innovation dynamics. Fuzzy translation of structural configuration of innovation system to functional performance.

4.4.1 Remaining blind spots

We have stated that innovation processes are dynamic, multi-level and multi dimensional processes and that theoretical frameworks should incorporate these notions. Three frameworks are found to do this to a certain extent. That is, more than linear models they provide guidance to study processes over time. However, if we would lay the three frameworks over the stylized notions of innovation processes a few blind spots can still be find. These blind spots will serve as impetus for the first contribution of this thesis: the introduction of an analytical framework from the fields of design studies that helps in analyzing systemic innovation processes. The following blindspots can be found:

1. Pure descriptive accounts of micro-level behavior are related to emergent system functions (on the niche level). This provides little help in cross-comparing cases.
2. Micro-level behavior is focused on manipulating the structural parameters to achieve functions, not on manipulating functions directly.
3. Cumulative causation between functions always happens through individual behavior, as it is the motor of change. Functions, contrary to individual actors, have no agency. Structural parameters therefore act as translation devices. These should be technological, social and institutional. Manipulating innovation processes is thus substantive design, network design and institution design.

4.4.2 P-S-I framework as linking pin

The P-S-I framework originates from work done at the Engineering Design Research Center at Carnegie Mellon University (Subrahmanian et al., 2011) based on the recognition that design is inherently a social process (Bucciarelli, 1984). Engelmore and Tenenbaum (1990) uncovered that the average designer spends about 15% on technical computations and 85% on coordination and negotiation through telephone calls and meetings. The model first defines design as the "purposeful activity aimed at creating a product or process that changes an environment or organization" (Reich et al., 1996). This activity is social in that design knowledge is not possessed by one homogenous consciousness (i.e. the sole designer) but is shared across a multitude of designing entities. Furthermore the model is build on the premise that design is itself an evolutionary process where artifact description, gathered information and the design process are coevolutionary linked (Westerberg et al., 1997). Design is thus highly dynamic (Subrahmanian et al., 2003). Originally the framework stated that design can be conceptualized using two distinct spaces: the product space and the social space (Subrahmanian et al., 2011) but later added an institutional space to the framework (Meijer et al., 2014) to do justice the dynamics of rules during design processes.

The core idea of the framework we borrow for our research on sociotechnical transitions is that decision making on transitions is inherently a multi-actor design task where sociotechnical systems are designed (an artifact is manipulated) in an environment of multiple actors with different perspectives, world views and languages and where different institutions state the rules of the game. In addition, we say that the design process can be conceptualized as a dynamic interaction between a product space, a social space and an institutional space. In Table 4.5 we briefly elaborate on the three spaces and their dimensions.

Table 4.5: P-S-I framework to study dynamic design processes

Space	Dimension	Explanation
Product (P)	Structural complexity	Interrelatedness of system elements
	Disciplinary complexity	Amount of different disciplines involved (models, languages, vocabularies)
	Knowledge availability	Completeness of the 'theory of the artifact'
Social (S)	Inclusion	Permeability of social space, ease of entering and leaving.
	Number of perspectives	Number of points of view that are critical in discerning the product definition and solution from conception to implementation
	Capabilities and Skills	The abilities needed to bring a solution from conception to implementation
Institutional (I)	Ties	Strength of linkages between actors within the social space
	Institutional Structure	Overall structure of the linkages (market to hierarchy)
	Knowledge accessibility	Ease by which knowledge can be incorporated into the social space.

The P-S-I framework stands right in the middle between theories on complex artifacts and theories on complex decision-making. The framework assumes that the transformation of sociotechnical systems is a design activity where dynamics in the complex technological artifact have a bidirectional influence on the network of actors that are involved in the design process. Similar notions can be found in Geels' model of the bidirectional influences between the sociotechnical system, the human actors and the rules and institutions (Geels, 2004) but this framework allows the scholar to more specifically study on what dimensions these three concepts change and how the one impacts the other. Other approaches that have had a resembling focus are for instance the interaction between organization and technology (Leonard-Barton, 1988), actor-networks in niche formation (Caniels and Romijn, 2008; Hermans et al., 2013), project and system interfaces (Geyer and Davies, 2000) and politics in innovation processes (Hardy et al., 2005). We contest however that all these models have always only linked two of the three needed models in describing multi-actor transition processes of complex sociotechnical systems.

4.5 Conclusion

As stated in the introduction the study of innovation needs to do justice to both the macro and micro level patterns in processes as well as tell the technical, social and sociotechnical story of innovating in systems like the railways. With the frameworks we have discussed so far we believe that we are able to cover the most important facets of innovation processes:

1. The RAFWS framework is a suitable framework for uncovering typical agent-level strategies for coping with complexity (Micro)
2. The NK-model is able to portray the technical part of the innovation journey (Macro)
3. The MLP-model is able to portray the macro-level dynamics of systemic innovation processes (Macro)
4. The TIS-model allows us to study the functions of an innovation system (Meso)
5. The PSI-model allows us to study the structure of an innovation system (Meso)

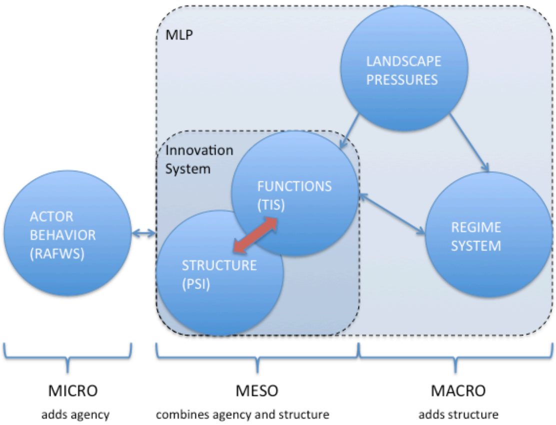


Figure 4.7: Multiple frameworks in relation to each other when used for studying macro and micro level phenomena in systemic innovation processes

In Figure 4.7 we provide an overview of the frameworks and how they relate to each other. We have excluded the NK-model because the model does not deal with the sociotechnical reality of innovation processes but much more is a way to conceptualize these processes using a specific nomenclature specifically focused on technological change. Hence, its exclusion does not mean that the model will not be used to make sense of our case studies later on this thesis.

In Chapter 3 we extracted from our interviews a plethora of notions that characterize innovation processes in the Dutch railway sector. With the combination of models in Figure 4.7 we believe that we are able to capture most of these characterizations in our case studies. Structure and agency, macro and micro, goal-orientation and emergent phenomena are all included. We believe such a combination of many models is especially important if one intends to conduct a single case study and to uncover interesting mechanisms at work from the ground up. Beforehand, we have no indication what these mechanisms may be nor are we able to uncover these without any theoretical or analytical framework. By including many of such frameworks we will be better able to creatively interpret empirical data and distill interesting notions. This comes however to the detriment of a certain rigor. The consequent use of one theoretical framework, even if a framework is not able to capture all phenomena of a process under study, helps in better disciplining empirical research and supports cross-comparison between different processes. Therefore, as we will show in Chapter 5, we will use on theoretical framework for our multiple case study and use all of the frameworks in Figure 4.7 for theory building in our single case study.

5 Research Methodology

The previous two chapters dealt with the design science approach to gaming and theoretical frameworks on innovation processes. The fundament on which this thesis builds is that any fruitful design of a game which lends its legitimacy from an environment that uses this game, can only occur when one knows what design options a game can have and what the context-in-use needs from a game. By looking at how the design of games relate to environmental needs, such as for instance the need to build a niche or to allow for creative exploration of designs, this thesis however remains mostly an design science endeavor: it incorporates the notion that subjectively stated assessment criteria guide the design of an artifact (in this case gaming simulation). This thesis will however not assert how one should design a game but merely intends to uncover what game design decisions are related to what outcomes in the innovation process in which such a game is employed. Subsequently the research methodology that this thesis applies is mostly stemming from the design sciences but uses theoretical lenses from the analytical sciences as well. Being a social science topic we use a methodology that does justice to the erratic nature of the research object caused by the large role played by human agency. In this chapter we will elaborate on the underlying ontology and epistemology of our study, the cases we pick for our study as well as the research methods we apply.

Whereas orthodox social sciences mainly deal with relating variables in a web of linear causal relations, this research takes on a process perspective. That is, this thesis will focus on building a narrative of innovation processes, chains of events, that explain how innovation processes in the railway sector emerge, change and terminate. By doing so, it departs from 'usual' management science literature in two ways: firstly, the thesis will mainly focus on inductively building theory that is context-specific, rather than testing theory that is context-free. Secondly, employing the case study method for this purpose, the thesis will focus on processes as event-sequences rather than the process by which variables are related. We do this based on the assumption that chaos theory and complexity theory are suitable ontological frameworks to understand and improve organizational processes (Stacey, 1995; Thietart and Forgues, 1995; Rotmans and Loorbach, 2009) and that narrative styles of explanation are more suitable for describing these complex processes than variance based approaches (Tsoukas and Hatch, 2001). Because of this departure from usual ways of studying social science phenomena, we briefly introduce the underlying ontological and epistemological considerations before delving deeper in the specific methodology this thesis applies.

5.1 Ontology

This dissertation is about two social science phenomena: innovation and interventions using gaming simulation. Although different topics, we address both from a specific ontological stance. That is, the phenomena that this thesis tries to describe and understand are social phenomena in the first place, and hence rife with chaos. Chaos is the seemingly random behavior of fully deterministic systems.

Chaos theory is relevant for this dissertation in several ways. Firstly, it supposes *practical* unpredictability for *theoretically* deterministic causal systems. Chaos arises when three variables are highly interrelated and dynamically correlate with each other over time. Due to this process the behavior of this causal system is highly sensitive to the initial values of the three variables. Secondly, for those interested in interventions into systems, such as organizational redesigns, training of personnel or some technological innovation, chaos theory shows that the effect of any intervention is potentially highly dependent on the history of the system as well as the specific configuration of that system when the intervention is implemented. The claim: 'this intervention worked' then becomes highly susceptible to internal validity threats.

The link with this dissertation is built exactly on these two relevance points mentioned in the previous paragraph. For this let us take the topic of the dissertation. "systemic innovation processes in railway systems and the role of gaming simulation". As we have shown in the previous two chapters, innovation processes are highly chaotic because of a constant interaction between the innovation artifact, those who design it, the rules they use, and the broader context. As such, there is no one innovation process and any two processes that start from almost the same initial conditions are likely to end up showing divergent behavior. Endogenously, the causal system that describes this innovation process for every cycle of mutual influence exacerbates even the tiniest of differences in the beginning. To say: in case A we saw this type of process and this process will happen for any other case that closely resembles case A, becomes highly problematic. In addition, this thesis wishes to shed light on the applicability and value of gaming simulation. As interventions, which employments of gaming simulation are, we must be wary of immediately concluding that the use of gaming simulation led to some consequences. Internal validity threats could come in the form of consequences not being the result of using gaming simulation but rather as a result from the historical processes already existent in the system. External validity threats could mean that we infer, unduly, that these consequences can be found anywhere anytime. Chaos theory could point us as researchers to the fact the consequences are potentially only there in this specific case, in this specific time, and under these specific consequences.

Furthermore as have demonstrated, railway systems can be classified as complex adaptive systems. Change within these systems is thus an emergent phenomena on the macro-level caused by the many interactions of multiple adaptable agents on a micro-level. Hence innovation processes consist of multiple levels that impact each other but to some extent operate independent from each other.

5.2 Epistemology

Usual variance based approaches simply assume linearity, even when there is none. Process based approaches however acknowledge emergence, non-linearity and chaos. In essence both perspectives differ in how they view the real world: either one consisting of stable substances only changing in their position in space and time or a world full of dynamic processes (Rescher, 1996). Respectively these ontological stances give way to either variance-based epistemologies or process-based epistemologies (Mohr, 1982). Within the variance-based epistemologies however processes can also be part of the research focus (Van de Ven, 1992). In those cases processes are either understood as something that can be translated into a variable (such as the speed of an innovation process) or the process of intermediating variables between the independent and dependent variables. We therefore distinguish three approaches and their related methodologies in Table 5.1:

Table 5.1. Different epistemologies and methodologies (Based on Van de Ven, 1992)

Approach	Focal causal relation	Methodologies
Variance-based content	A leads to B	Experiments, Linear regression analysis
Variance-based process	How A leads to B	Case studies, process tracing
Process-based	Sequence of events that describe how things change over time	Case studies, event-sequence analysis, narrative analysis, grounded theory

At first sight, the notion of chaos and complexity completely renders any scholarly activities useless. It might seem that no generalizable statement can be made on any system that has some chaotic properties. For instance, how are we able to fully model the weather if indeed the clap of a wing of a butterfly can cause a tornado 10.000 miles away? Does the meteorologist really have to incorporate butterflies into his model to be able to tell people whether or not tomorrow it will rain? Luckily, chaotic and complex systems are not random systems. That is, we are able to catch some of its properties with the models we use to explain and predict the behavior of these systems. These systems might have upper and lower bounds, attraction points, tipping points, and recurring patterns. Furthermore, different levels might in themselves be better describable in more variance-based terms. Back to our example, in the winter we tend to have cold weather and in the summer the weather tends to be much warmer. Although chaotic, we can most certainly say that tomorrow the temperature, anywhere on the world, will not exceed 70 degrees centigrade, regardless of a butterfly clapping its wings or not.

Returning to Poole and Van de Ven (1989) a good process model always has two complementary components: "*The global (macro, long-run) model depicts the over- all course of development of an innovation and its influences, while the local (micro, short-run) model depicts the immediate action processes that create short-run developmental patterns*" (p. 643). Such a process model would do justice to both the local immediate activities of agents in a complex adaptive system and the more emergent systemic patterns. Thus, we can describe innovation processes occurring in complex chaotic systems if we focus our analysis on both the overall patterns and the underlying mechanisms by which actor behavior caused these patterns. As Geels (2011) states: *"The focus on patterns and mechanisms may enable transitions research to articulate an epistemological middle way between on the one hand the search for laws and statistical correlations between variables (as in mainstream social science), and on the other hand an emphasis on complexity, contingency, fluidity, untidiness and ambiguity (as in constructivist micro-studies)"*

5.3 Methodology

Process research deals with things that emerge, develop, grow and terminate over time and tries to explain the how and why of these processes (Langley et al., 2013). Two important factors distinguish process research from variance-based research: the focal unit is the event and temporal progression is the core of the explanation. Events are changes in each of the constructs of the conceptual model by which the analysis is done (Garud and Van de Ven, 2002). Temporal progression hence means the ordering of events in event-sequences (Langley, 1999; Pettigrew et al., 2001). Although we first and foremost take on a process-based approach, we still intend to measure some variables. Hence, our approach is synthetic (Langley, 1999), bridging the process-based event approaches and the process-based variance approaches.

This study will first and foremost use a qualitative research methodology. By doing so, we remain close to the reality by describing, in qualitative terms, what is happening and why. Rather than constantly measuring in variables the state of affairs during an innovation process, we use a more narrative-style, portraying a sequence of events as the main focal point of our study. This helps in staying close to what reality we try to describe (ontology) and in what ways we can know this reality (epistemology): a reality that involves the highly contextual activity of design.

We use case studies as our main methodology by which we intend to answer our research questions. As the previous two chapters have stipulated systemic innovation processes in embedded systems, such as the railways are, can be characterized by serendipity and the pervasive impact of context. Case studies are perfect for investigating phenomena within its real-life context (Yin, 2013). For studying innovation, case studies are becoming more and more popular because of this very reason (see for instance Verganti, 2008). Innovation processes are open and hence demand the researcher to give great consideration to the context in which these processes are embedded (Pettigrew, 1990). Also case studies enable the introduction of the temporal and dynamic component inherent to processes in general and in models on the use of gaming simulation in particular (Klabbers, 2009). They are also suitable for theory generation rather than theory testing (Eisenhardt, 1989).

Since we are interested in both macro-level patterns and micro-level mechanisms and they both cannot be understood in complete isolation, this poses a problem for our research methodology. Ideally, one theorizes on both simultaneously, and to some extent this has been done in this research. However, one always needs to fit into the sequential order of a written text. Although the theorizing on both the macro and the micro has been done simultaneously to some extent, we feel that the order of macro to micro in both our methodology and chapter ordering was unavoidable due to practical reasons as well as defensible for methodological reasons. If we compare the two empirical chapters on innovation processes, dealing either with macro-level patterns (Chapter 6) and micro-level mechanisms (Chapter 7) they differ especially in their level of analysis as can be seen in Figure 5.1:

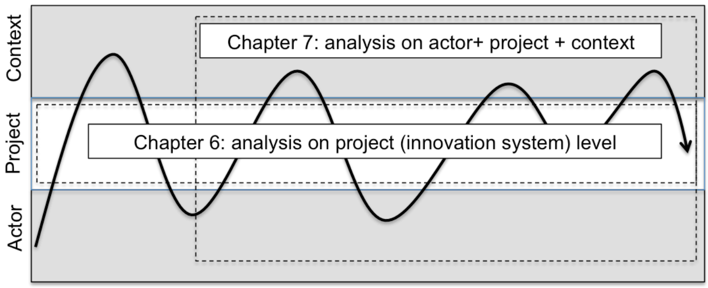


Figure 5.1: Unit of analysis of chapters 6 and 7 compared.

To find something 'interesting' one always needs to compare something. Hence we first seek for macro-level patterns that are unique to systemic innovation processes by comparing one case with two other cases. Then we use the macro-level pattern as a disciplining force on the more inductive approach used to uncover the mechanisms found in Chapter 7. Because to find these mechanisms one needs to take a broader view then the sole project (because driving mechanisms might find their origin from the context or from individual actor behavior) we need a plethora of theoretical frameworks to make sense of an otherwise messy unit of analysis. Also, both chapters differ in that the first uses a multiple case study and the latter a single case study.

Chapter 6 namely starts with a multiple case study in which we try to uncover the distinction between systemic innovation processes and other non-systemic innovation processes. As the research question stated we are interested in macro-level patterns in innovation processes and how the makeup of the innovation (being systemic or not) influences these patterns. Multiple case studies do however need a structure analytical framework beforehand for else no comparison is possible (Aboen et al., 2012). Also for a good comparison we need similar delineations of our object of analysis, so we keep our focus solely on the project-level and leave the context out of our analysis. Although we have stated context does matter, for comparison reasons this chapter will keep a limited scope. We chose the PSI framework as the analytical framework by which we compare three cases. We do this for the following reasons: firstly, the framework is relatively theory-free: it does not state how innovation processes take place but merely provides the researcher a framework to assess the structural features of an innovation system. Secondly, the PSI-framework is entirely focused on the sole innovation project and therefore allows for delineation of the object of study. Especially the MLP-framework has been criticized for delineation issues (Smith et al., 2010). Thirdly, given that the framework solely focuses on structure and not on functions, which the TIS framework does, it is less prone to post-hoc rationalization and hindsight bias (see Bizzi and Langley, 2012), where respondents might portray their historical actions as retroactively intended to create certain needed functions, such as market formation or learning. Also, given that in all three cases, which will be elaborated on in the next paragraph, have reached their implementation stage, we can safely assume that most of the functions of the innovation system have been performed and that a comparison using functions and not structure renders no interesting results.

To further allow for comparison we have to fit the empirical data in some template because otherwise context specific case-unique events might render a comparison invalid. This especially concerns the timing of these processes where the exact timing of changes in the structural dimensions of P, S and I are not so interesting. For instance, if the complexity of the artifact in project A becomes higher in year 3 and in project B in year 5 what does this tell us? And what if project A was a ten-year project and project B a five-year project? Therefore we use the phase-model of (Rotmans et al., 2001) that is used in systemic innovation process instead of a true temporal ordering of events. This somewhat linearizes our analysis, from which we originally tried to refrain ourselves, but we make this concession in order to be able to compare cases. The resulting scheme for analysis then looks as shown in Table 5.2:

Table 5.2: scheme for analysis for multiple case studies on macro-level patterns

	P	S	I
Pre-development	Dynamics in product makeup before take-off	Dynamics in social space before take-off	Dynamics in institutional space before take-off
Take-off	Dynamics in product makeup during tipping point in process	Dynamics in social space during tipping point in process	Dynamics in institutional space during tipping point in process
Acceleration	Dynamics in product makeup after tipping point	Dynamics in social space after tipping point	Dynamics in institutional space after tipping point
Stabilization	Dynamics in product makeup during phase where innovation settles in the regime system	Dynamics in social space during phase where innovation settles in the regime system	Dynamics in institutional space during phase where innovation settles in the regime system

In Chapter 7 we take into account the context and individual human behavior to try to explain the pattern found in Chapter 6. For this we take a much broader view and use a single case study: one that involves a systemic innovation. A single case study is perfectly suited for continuously juxtaposing the empirical case and theoretical frameworks (Dubois and Gadde, 2002). Given that no single theoretical framework is needed beforehand as well as a strict delineation of the unit of analysis, single case studies are also well suited for taking into account a potentially limitless context. Furthermore we broaden the scope in time as well. For comparison purposes we limited ourselves to the project-level in Chapter 6, prohibiting us from studying the evolution of the innovation idea itself even before it had been put into a project. Now the single case study allows us to do so, since we have acknowledged in our ontology and epistemology that history matters and that historical processes potentially echo in the processes we currently observe.

The juxtaposition of theory and the single empirical case will inherently involve a bidirectional influence. Although this means that the analysis will not be theory-free or done without any predetermined theoretical lens, the use of more than one theoretical framework will allow us to use a method that is somewhat similar to a grounded theory approach (Strauss and Gorbin, 1990). We partially transcribe the interviews into general sentences that describe what is being said. From the list of sentences we compile codes that summarize a set of sentences that have been put forward by the respondent in sequential ordering. Then we try to relate different codes together and see what categories emerge, termed axial coding. From these categories we try to distinguish potential mechanisms that explain why we have found the unique macro-level pattern in Chapter 6 that solely applies to systemic innovation processes. When we have found these mechanisms we bring the aforementioned theoretical frameworks to the analysis and look where our findings complement current frameworks and where they contradict each other. This way, the frameworks both are there to make sense of what we have found as well as tell us where the found mechanisms provide new insight into systemic innovation processes in networked infrastructures.

In Chapter 8 we go back to a multiple case study. Firstly, we distinctly look at each of the cases individually and assess to what extent they impacted the macro-level pattern found in Chapter 6 and through what mechanisms from Chapter 7 it did so, if at all. It is at this place the Design-in-the-Large and the Design-in-the-Small are combined. Different from the analysis of gaming in Chapter 3 we use no theoretical or analytical framework to assess the relationship between gaming simulation, innovation mechanisms and innovation patterns but let the empirical data speak for itself. This chapter is the core of this thesis in that it bridges gaming methodology, innovation theory and innovation practice and we have shown how this field is still relatively underexplored. Therefore we wish to inductively come to a set of propositions on the value of gaming simulation for innovation processes and through what design choices game designers and game professionals can realize this value.

As for almost any gaming simulation exercise, the exercise of this thesis will also end with a debriefing. In Chapter 9 we devote considerable effort to constructing a framework to properly debrief games for innovation processes. In this chapter we wish to build a framework to capture the true value of gaming and to connect gaming as a black-box exercise and the outside world in which this exercise is conducted. The framework is built on the premises, propositions and hypotheses put forward in the previous chapters and our speculation on how to both reap the benefits of the promises of gaming and to alleviate some of its shortcomings. This final chapter is therefore less methodologically grounded than the other ones and will cautiously provide a framework for debriefing with the caveat that further research, even practical testing of the framework, is needed.

5.4 Case selection

The general methodology by which this thesis aims to answer the proposed research question is the case study, being either a single or multiple case study. For Chapter 8 (innovation practice and gaming methodology) we are of course limited to the games that we have designed and executed in the railway sector. From this limited number we exclude those games whose initial purpose was to create effects other than the generation or testing of hypotheses because these games have been sufficiently discussed in other literature as well as they do not provide interesting lessons for innovation theory, which mainly deals with experimentation. This is not to say that games for learning of games for policy making cannot have any value for innovation processes, but these games fall outside the direct scope of this thesis. From the remaining range of games for research and design we will not exclude more games because we intend to inductively search for the theoretical value of games and the practical value if we take into account its embedding in ongoing processes. Note here that from Chapter 3 we have concluded that the value is likely starkly different than simply providing valid claims about pre-generated hypothesis. It is this contrast between initial purpose and potential value that we are after in the multiple case study.

The multiple case study in Chapter 6 strives to find patterns that make systemic innovation processes unique. We therefore look at how other innovation processes unfold given the PSI-framework as our analytical framework. In the introduction we have mentioned the distinction between radical and systemic innovations. Systemic innovations are innovations that are characterized by two dimensions: the interdependence of innovation-internal elements (they are collections of otherwise incremental yet dysfunctional improvements, their systemness is not guaranteed on the outset) and their overlap with the legacy system (they not only change the makeup of the system but also the way the system works). Given these two dimensions we have a 2x2 typology in Table 5.3 to distinguish the following innovations:

Table 5.3. A 2x2 typology of different innovations

	Low external overlap	High external overlap
Low internal systemness	-	Systemic innovation
High internal systemness	Radical innovations	Civil engineering projects such as a railway bridge or tunnel

It is in the relation between internal systemness and external overlap that we expect to find interesting distinctions. High external overlap will result in regime pressures (since innovation elements not only serve functionalities of the innovation but also for the current system) and low internal systemness will result in innovation elements more easily being influenced by these regime pressures. To add to this analysis we provide an anchor in the form of a case where the innovation scores high on both dimensions (a tunnel of which the internal systemness is high: you cannot built a half tunnel). We have found three cases that fall in this 2x2. We note that so far we could not think of a case where the internal systemness and the external overlap are low, but incremental innovations might fall under this label.

Radical innovation
In the UK they are currently overhauling the outdated traffic control processes and installing a new traffic management system (TMS). Although initially the architecture of the innovation was different in the end the TMS is bought off-the-shelf and additional interfaces were designed to make it overlap less with other ongoing innovations as well as the legacy system.

Civil engineering project
In Delft they are building a railway tunnel under the historical heart of the city. The project Spoorzone Delft is tasked with designing and commissioning this railway tunnel. Given that it has to fit the existing system (its technical makeup as well as all the nation-wide procedures for operating it), the external overlap is high.

Systemic innovation
Since the Japanization of the Dutch railways is the most prominent systemic change we chose the DSSU project, which is involved in bringing about elements of this Japanization in the rebuilding of the infrastructure and railway station of Utrecht Central. Its internal systemness is low since all measures (separation of corridors, signal optimization, removal of switches) could technically be done separately. Even if half of these measures are implemented, they could provide some value. This is in stark contrast to a half built TMS or tunnel.

In Chapter 7 we focus more deeply on the DSSU case and distant our analysis from the sole project. Rather this single case study will study the evolution of the entire 'Japanization' idea and how it finally got implemented through the DSSU project. Given that the case study no longer focuses on an easily demarcated project results in an analysis issue: what is still part of the case and what is not? For this reason, what is considered the case is partly discovered during the case study itself. Additional information, events and respondents are studied as long as they increase the amount of insight into our main focus of this case: the mechanisms underlying the pattern we found in Chapter 6. This is theoretical saturation and is a suitable way of defining the case study after the fact.

5.5 Methods

Sources for our narrative of different innovation journeys and the intervention of gaming simulation will range from interviews with involved project managers, observational research during gaming simulation sessions as well as general project meetings and written documents such as project reports. This range of sources allows us to depict innovation processes from many angles.

For Chapter 8 we relied solely on our own involvement in designing and executing gaming simulations for the railway sector thus the focal method was participant observation. In Chapter 6 we use a plethora of methods to retrieve data on dynamics in the P, S and I spaces over time and these are, for practical reasons, different per case. We were involved in commissioning meetings for the tunnel projects and able to observe the monthly discussion over a period of four years. For the TMS project in the UK we held 10 interviews with involved stakeholders at Network Rail, the ProRail-equivalent for the British railway sector. These stakeholders ranged from general managers to project managers and ergonomics subject matter experts. We came across this project at a certain point in time where most of the interesting dynamics had already occurred, forcing us to rely solely on retrospective accounts of interview respondents. For the DSSU case we were asked to help organize a series of three meetings to discuss pressing issues related to implementing the project. These meetings in themselves already led to valuable insights and from these meetings resulted an additional five interviews with key players in the process.

The more in-depth case study in Chapter 7 encompassed an additional 25 interviews to the five already conducted for the DSSU case from the previous chapter. Table 5.4 shows the backgrounds of each of the interview respondents. We added to these interviews a range of project documents and progress reports as well as research documents and presentations on certain pressing issues. We adopted a snowball method where each respondent could point to additional interesting respondents or documents. We stopped interviewing as soon as we saw that no new interesting phenomena could be uncovered.

Table 5.4. Overview of interview respondents (one double interview with innovation department member, hence 24 respondents)

	Routing / Time-tabling / Capacity planning	Safety	Traffic control	Project / Innovation	Representative role for respective organization
ProRail	2	3	1	5	
Dutch Railways (NS Reizigers)	5	3			
Nedtrain					1
NS Hispeed					1
Contractor				1	
External					2

We provide a summary of the methodologies, methods and the role of theory in our analysis in Table 5.5:

Table 5.5 Overview of methodology per chapter

Chapter	Methodology	Supporting role of theory	Methods	Goal
Chapter 6 Theory vs innovation	Multiple case studies	Theory as an analytical framework	Participant observation, Interviews, Documentation	Generate theory on relation between type of innovation artifact and macro-level patterns
Chapter 7 Innovation	Single Case study	Theoretical framework to explain empirical findings after the fact	Interviews, Documentation	Generate theory about micro-level mechanisms specific to systemic innovation processes
Chapter 8 Innovation and gaming	Multiple case studies	No theory	Participant observation	Generate theory on elements of gaming simulation that impact pattern through mechanisms
Chapter 9 Debriefing	Design science approach	No pre-existing theory, save for theory emerging from our own analysis	Participant observation	Design a debriefing framework built on what is learned so far

5.6 Conclusion

In the theoretical chapter on innovation processes we claimed that systemic innovation processes are highly multi-faceted and therefore need the use of many different theoretical frameworks and analytical frameworks to fully understand all the relevant phenomena at play during such a process. Similarly, we conclude here that such an eclectic design science approach is also needed in regard to the use of a research methodology. Our intention is to uncover relevant mechanisms in a single case study but for this purpose we need to first determine what relevant is through a multiple case study. Also since we do not know beforehand where such relevant mechanisms can be retrieved empirically, we use a broad set of data collection methods, most of these opted for as the research underlying this thesis went on. Disregarding obvious practical concerns regarding the availability of data and respondents, inherent to doing research on processes not instigated for the benefit of the researcher, the variety in different data collection methods corresponds to the many ways by which we expect interesting phenomena to manifest itself during a systemic innovation process.

6 The Uniqueness of Systemic Innovations: identifying macro patterns in innovation processes from three case studies

This chapter seeks to find a unique macro level pattern for a systemic innovation process that serves as a disciplining force on the exploratory study in the subsequent chapter. In Chapter 7 a single case will be studied to see what micro-level mechanisms are at work when an otherwise assumed inert railway system is confronted with a systemic innovation. To find the relevant micro-level mechanisms, we first need to know what macro level pattern is both unique and problematic to systemic innovation processes. Otherwise Chapter 7 will result in micro-level mechanisms that are indeed at play in real life but do not provide any knowledgeable input to Chapter 8, where we will uncover the ways by which gaming simulation can and cannot support systemic innovation processes. Hence this chapter studies three different types of innovation processes using the PSI framework and asks the question:

To what extent are the structural dynamics in the product, social, and institutional spaces over time qualitatively different for systemic innovations in the railway sector than for other types of innovations and what is this qualitative difference?

To answer this question, this chapter looks at the macro-level patterns of three different projects dealing with different kinds of innovations, from incremental to systemic. The analysis is purely descriptive in that it wishes to show, but not fully explain, the differences in dynamics in the innovation artifact, the involved actors and the rules and norms they use over the lifetime of the project. Far more so than in the subsequent chapter our analysis is focused in that it analyzes solely the project and its immediate stakeholders over a bounded period of time. Furthermore we use a structured analytical framework to assess the difference between the projects by using the PSI framework. Doing so, we trade in depth with breadth since we do not so much look at the detailed underlying causal factors that explain why the projects are different but more look at what makes them different in the first place.

To find this difference, this chapter firstly applies the PSI framework to each of the three case studies individually. Using a range of data collection methods, as we have elaborated on in the previous chapter, we arrange our data according to changes in the makeup of the technological artifact (P), the involved actors (S), and the institutional setting (I) over four different phases of an innovation process. The comparison of these changes between the three different projects is then done in the analysis paragraph of this chapter. Here we look for each of the three spaces (P,S and I) how these evolved differently between the three case studies. A final synthesis of all the findings is given at the end of that paragraph, where the combined dynamics of the P,S and I spaces are compared. This chapter ends with a conclusion of the main findings and how these findings can benefit our analysis in Chapter 7 and existing theoretical frameworks. We firstly start with a brief recap of our methodological considerations and introduce the three cases.

6.1 Analytical framework and case selection

In Chapter 2 we have elaborated on the many theoretical and analytical frameworks available to the systemic innovation scholar. As we have stated, the Multi-level Perspective of Geels (2002) is a framework designed to make sense of these transition processes specifically. Other frameworks, such as the NK framework (Kauffman, 1993) and the Technological Innovation Systems framework (Carlsson and Stankiewicz, 1991) are suitable to study these systemic change processes as well but not exclusively. They can also be used to analyze, understand and improve more incremental innovation processes, or to make sense of innovation and technological change in general. Although the MLP might then be considered the most suitable to study systemic innovation, we contest that even when using solely this framework, one might forgo on seeing interesting dynamics in the innovation process. This is because we fear the MLP falls short on two aspects of studying innovation processes. Firstly, the MLP presupposes certain dynamic patterns in these processes, such as convergence of actor behaviors towards one common goal, the creation of one dominant design, and increasing stability in the process. Logically, one would at least feel that when such processes happen as an innovation moves out of a niche and into 'the real world', these processes would be counteracted by increased pressures from the regime. In that sense, in the MLP the regime seems to be a passive actor, merely receiving the new systemic innovation. Secondly, we feel that although we stated that autonomous innovations and systemic innovations are qualitatively different, their effects on innovation processes might not be qualitatively different. Although autonomous innovations inherently do not involve traversing a valley in the fitness landscape (and thus see no temporarily decrease in performance) they still might be perceived as occurring along the same route as the MLP exclusively awards to systemic innovation processes. Autonomous innovations, although a logical point forward for the system, have to compete with other autonomous innovations for resources and managerial attention. In that sense, there is a logical argument to be made for these smaller scale changes also to originate from niches and to be impacted by regime pressures. Then the dynamics the MLP assumes solely for systemic changes might be applicable to autonomous changes as well, albeit on a smaller scale. Inherently systemic innovation processes are more complex, involve more stakeholders and occur over longer periods of time but these differences are of scale and not qualitatively different. If the differences are merely quantitative, then what such processes demand from gaming simulation are no different in quality than what incremental innovation processes or simple improvement projects would demand from gaming simulation. Hence we feel that the MLP will not suffice to distinguish a systemic innovation process from other innovation processes.

In the previous chapter we therefore stated that in order to find a unique pattern for a systemic innovation process, we need an analytical framework that is more theory-free. We chose the PSI framework because this framework merely gives the innovation scholar a set of 9 parameters to describe an innovation process. The PSI framework does this without presupposing how any innovation will or has to look like in real life.

To find an interesting pattern we looked for cases that involved different innovation artifacts. The difference between these artifacts, as we mentioned in the methodological chapter, can be found in their architecture (the way the elements that make up the innovation are connected) and their overlap with elements of the regime system (the extent to which elements are both part of the innovation and the legacy system). We note here that these differences are found after the fact as we have looked at innovation processes in retrospect. This means that for the processes under study the initial innovation might be different with respect to its internal architecture or its overlap with the system it is trying to replace or invade.

Spoorzone Delft (SZD)

The first case involves the least innovative artifact: a railway tunnel under the historic heart of the city of Delft. Although not an innovation, it is a new artifact to most of the involved stakeholders and it also involves the design, testing and implementation of an artifact, similar to designing an innovation. We include this case because it bridges two distinct innovations (radical and systemic) in that it resembles radical innovations because of the interdependence of its elements and in that it resembles systemic innovations because it overlaps much with the existing system. This overlap is there because the tunnel replaces an existing line and many of the elements that make up a tunnel are also part of the existing national network. Signaling, operator procedures, safety procedures, railway station layouts are not solely elements that the directly involved stakeholder can design at will. These elements also fall under the responsibility of stakeholders that are part of the existing system. Hence what might be optimal for the isolated tunnel, might not be optimal from the viewpoint of the existing system.

Traffic Management System (TMS)

This case involves an overhaul of the traffic operations at Network Rail, the U.K. equivalent of ProRail. The core of this overhaul is the introduction of traffic management and the implementation of a traffic management system. Traffic management is the intelligent control of traffic if traffic is no longer able to function according to the predesigned timetable. The traffic management system will provide human operators with technological support by automating otherwise manual processes and by providing additional intelligence and forecasting methods. The system itself is similar to the tunnel project in that the elements that make up the system are highly interdependent. For both projects hold that one cannot implement it for 50% since then the innovation artifact will not function. However, regarding the overlap the system very much overlays the existing system rather than replaces it, as the tunnel does. It is added-on to the legacy system because of the installation of interfaces between the innovation and the existing system.

Doorstroomstation Utrecht (DSSU)
The rebuilding of Utrecht central station according to Japanese principles is then very much a mirror-image of the traffic management case. Since the innovation is much more a collection of elements that can be seen loosely from each other and also implemented loosely from each other, the interdependence of innovation elements is much lower. Furthermore the elements that this innovation encompasses are overlapping highly with the existing system because it does not involve adding-on something but much more reconfiguring the layout and working of the already existing infrastructure and operations.

6.2 Data

For each case study we used a different data collection method because of practical concerns. To streamline the analysis later on in this chapter, we first arranged the data so that they tell a story of relevant events over time, from the initial invention of the innovation idea until its final implementation. For this purpose, we use the P-S-I framework to depict how the three different spaces evolve over four different phases during the process. Adhering to the MLP-framework we typify these phases in the process as a predevelopment phase, a take-off phase, an acceleration phase and a stabilization phase (Rotmans et al., 2001). This first empirical paragraph shows the results of this analysis on an individual case study basis. For each case we build a storyline of the innovation process and then translate this storyline into our analytical framework. In the subsequent paragraph we compare the individual case studies in order to analyze the differences and similarities between the three different innovation processes.

6.2.1 Network Rail

Of all railway sectors in Europe, the UK has by far been the most apt in privatizing the industry. Since 1993, Railtrack was created to manage the infrastructure on which many newly founded private train operating companies (TOCs) were given franchises to commercially exploit the tracks for a certain amount of time. Most notably in this case has been the steady decay of the infrastructure caused by Railtrack not knowing the quality of its assets. Factual asset management was contracted out to a service and renewal company. At the turn of the century several train crashes caused by metal fatigue led Railtrack to instigate a network-wide speed restriction. The resulting loss in economic performance eventually resulted in the organization going into administration. In 2002 Network Rail was formed as a successor with tighter control by the government. From those years on the UK railway sector gradually improved the conditions.

It is in this light that around 2006 some within the organization sought to improve the situation by looking at how other countries were managing their infrastructure. It appeared that one key difference was the much more developed traffic management systems employed. Traffic management is the intelligent control of traffic on a network by making sure that traffic controllers make those decisions that best help in adhering to a predetermined timetable. In its simplest form, traffic management comes down to a traffic controller rerouting a train from platform 1 to platform 2 if that helps in diminishing delays. The system then aids the traffic controller in his or her decision making by providing insights in decision consequences or by suggesting best practices during disruptions. The traffic management system can be automated to such an extent that for minimum delays no humans are involved in rerouting trains, as is the case in the Netherlands. In the UK, however, Network Rail has mainly relied on the manual operation of railway switches using human signalers located throughout the country. Still, within the broader organization the specific problems were not pregnant at the time. Therefore, as a respondent mentioned, the process really started with a solution rather than a problem.

Between the years 2006 and 2011 the product space can said to be very dynamic given the high exploratory nature of the project in that time frame. Many solutions were discussed, such as ERTMS and DAS, driver-advisory system advising train drivers on the optimal speed, and a yet to be designed traffic management system. The driving champion for this system then started the operations development project to make a business case for the TMS. Initially this system would be designed in-house as is usual for Network Rail, but later on to ease the implementation process the system would be purchased off-the-shelve.

The project took-off when the McNulty report, commissioned by the department for transport (DfT) and the office of rail regulation (ORR) was published in May 2011 and Network Rail was seeking for non-invasive improvements to the operations of the network for the next budgetary period (termed 'control period 5') (Department for Transport, 2011)**.** In the report McNulty demanded better utilization of government money on rail infrastructure, more efficiency in the sector as well as better cooperation. In combination with the upcoming control period 5 (the franchise agreement under which Network Rail works for the government), the organization had to find ways to show that they were ambitious and keen on meeting the challenges put forward by the report. Here network rail starts the Orbis project to digitalize most of the information channels at Network Rail by 2020 and cut costs. Although not directly stated, the sudden take-off of the TMS project at this time, after around 6 years of discussion is noteworthy.

At the end of 2012 the Ops Development board decides to hand over the TMS project to infrastructure delivery department to speed up the implementation. Respondents note that this switch from exploration to implementation was quite sudden. At the same time the ambition for TMS and its business case become more limited. From the many initial advantages only a 20% reduction in reactionary delay remains. Concurrently, TMS is separated from other projects that in conjunction could have delivered additional benefits such as higher capacity and lower costs. Also the project team decides to procure interfaces between the TMS and the interlocking and between TMS and the legacy system for data exchange. That way, the TMS system becomes modularized from other elements of the UK rail system.

This period also shows two interesting events. Firstly, the project decides to buy the traffic management system off-the-shelve. Secondly, in 2013 they decide to invite representatives of train operating companies to join the project, something new to the sector. It was the teams' explicit desire to make the TMS project user-led rather than engineer-led as was usual for the sector. In addition this period sees increased involvement of the unions as the TMS will have significant repercussions for the job roles of current traffic control staff.

Network Rail opted for an approach that loosely coupled the TM-system to the regime. In many instances has the social arena explicitly stated that they would like to keep the product space manageable by structurally decoupling it with other concurrent innovation processes. With buying the system off-the-shelve, they did not search the fitness landscape but rather cherry-picked solutions that originated in other European systems. To allow for adaptation in the first place, they weeded out suppliers that could not adhere to the functional requirements the organization made based on a concept of operations. This concept of operations, stating how traffic control roles and rules were to be applied in the future was leading in the further procurement of the TMS system. This has kept the project relatively manageable since during later phases in the process, the impact of the TM system on either other niches or the regime remained relatively stable and the market already solved the complex problem of designing the internal architecture of the system. As one respondent mentioned, this caused the forming of silos where different projects became separated over time and did not communicate any longer on finding possible synergies. Also, during prototyping using real operators, thereby bringing the innovation and the current system closer to each other, additional requirements arose about the functionality of the system. It proved that, due to high epistasis of the internal architecture of the TMS, this came either at significant costs or was not possible at all. As for now, the system is being implemented at several operator centers throughout the UK and with the overall operating strategy roll out spanning the next 30 years.

The gradual implementation process per operator center nationwide also shows an interesting feature: operator centers in the devolved regions can freely opt for early installment of the full TMS or use a phased approach where some of the features of TMS are already installed before full deployment. Although not fully deployed, the TMS that is then installed is skinned-down version but not a qualitatively different system serving different functionalities. The high momentum of the system as it is now however, according to some respondents, causes little flexibility in adapting to new unforeseen consequences or progressive insight. Also, it leaves the operating centers with little flexibility in adjusting the system to their specific needs. In Table 6.1 the overview of the interview data is presented.

Table 6.1. Network Rails' innovation process

Network Rail			
	Product	Social	Institutional
Predevelopment	Predevelopment since 2006, impetus given by benchmarks on other European infrastructure managers. Exploration (and expansion) of product space, studying many different products in other countries. Additional dynamics in disciplinary complexity and knowledge due to the seeking for problems that the TM-solution could solve. Different functionalities become part of the product space and their linkages with system elements are complex.	One product champion seeking funding for investing in a new TM system. Inclusion limited to those who are studying the principles of TM and working on the business case.	Institutional structure 'matures' after privatization and initial chaotic relations and conflicts. All actors involved are used to new setup. Weak ties between different organizations due to ever changing franchisees.
Take off	Product becomes Commercial-of-the-Shelf (CotS), most complexity thereby outsourced. Knowledge availability guaranteed by contracting proven suppliers. Structural complexity mostly revolves around linkages with concurrent niches and with the regime.	External landscape pressure to cut costs and increase punctuality especially by a government report by Lord McNulty and the beginning of control period 5. Project team installed, looking for potential suppliers of which three are invited to build prototypes. New arrangements lead to new actors entering the arena bringing capabilities needed to switch from exploration to implementation.	Special project team formed to study potential suppliers. Organizational rearrangements happening under the new national operating strategy (NOS) program. Traffic management becomes part of a wider program focusing on improving the operating performance.
Acceleration	Structural complexity decreased by focusing on one functionality (resilience of traffic control) rather than a plethora, e.g., capacity, costs, maintenance costs, etc.). Also little synergies sought with other concurrent niches and with train operating companies.	Representatives of operational staff join project team; TOC cooperation depends on local context and franchise length but is limited nonetheless. Inclusion limited due to modularization of separate niches. Involvement of labor unions to avoid operator resistance.	New design rules added: concept of operations should be leading and provide direction in the further design of the TM-system. Design of new governance and project management principles to cope with the newness of the project
Stabilization	Stable introduction to the regime (as for now). Innovation builds up momentum due to invested interests and internal structure. Only small changes are possible.	Stable social space, disciplines involved focusing on introducing a ready-made product into the organization.	No changes in institutional space (as for now)

6.2.2 Spoorzone Delft

Spoorzone Delft involved the building of a railway tunnel through the city center of Delft, including a new underground railway station. The interdependencies between innovation-internal elements were strong, as all elements were needed for the tunnel to function properly. Furthermore the overlap with the regime was profound in that the tunnel should allow for standard operating procedures regarding tunnel operation, traffic control and emergency protocols. Regime players from the incumbent system mandate these procedures to maintain interoperability over the entire national network. This lead to many of the design elements the involved stakeholders were considering were in fact also design elements considered by regime players. This created the dilemma between local optimality (for the tunnel) and network wide optimality.

Before the initial start of the project, the tunnel was mainly a political discussion. The 1990 rail21 project stipulated integral doubling of tracks along the Amsterdam-Rotterdam corridor. The city of Delft immediately felt that an elevated track through the heart of city with four instead of two tracks would be undesirable. Therefore, already in the 1990s discussion arose about building a tunnel with an underground railway station. In this phase, prior to the start in 2005, the discussion was mainly about financing the project and not so much on the technical specificities of the design of the tunnel. Hence, political dynamics (such as a switch to a more conservative-leaning cabinet in 2002) were the crucial dynamics at this particular stage of the project. The tunnel project was temporarily stalled after the then government decided to focus more on investing in car transport than in public transport. Only after the right amount of funding was secured, did the project more and more involve substantial design decisions. However, at this stage these decisions solely involved esthetical issues, such as the design of the station area and station hall.

In 2006 the city of delft founded the company OBS to start contracting outside parties for the design of the tunnel, the station and the station area. ProRail is in the lead for contracting parties for the design of the tunnel and in 2008 awards the contract to CrommeLijn, a consortium of three civil engineering companies. In the meantime, additional funding is allocated for making the design future-proof by already building a second tunnel. That way, the entire tunnel is able to accommodate four tracks even if after the first commissioning only two tracks are available for use. The reason for this is the start of the PHS program that intends to provide timetable-less train transport on the corridor between The Hague and Rotterdam.

For handling the project, project members adopted standard systems engineering principles to structure work processes mainly related to the technical parts of the tunnel. Firstly, this involved specifying all requirements beforehand, i.e. determining what procedures the tunnel should allow for. Secondly, this involved hierarchically structuring the design: from an overall grand design, to more detailed modules beneath it. This way, coordination was mainly realized using hierarchy where higher-level designs constrain the degrees of freedom for lower-level designs. This approach caused many perspectives and disciplines to be involved mainly in the beginning of the process. Later on, when designers could work on their specific part in modules, the work became more specialized.

Respondents acknowledged that this approach has its disadvantages. Faulty modularization creates interdependence in designs without coordination between design teams. This can cause modules to not fit properly when re-integrated in later stages of the process. It is cumbersome to design modules in such a way that after reintegration the overall artifact still adheres to the requirements specified beforehand. Modules create compartmentalization and each module might be impacted by its own eigendynamic as well as impacted by external pressures.

At SZD such phenomena were expected given experiences with similar projects. The building of the Betuwelijn and Hanzelijn, two large projects commissioned before the start of the SZD project, showed such problems as misfits between modules and between the overall artifact and the functionalities desired by end-users. The project team therefore put up an extensive testing regime. This involved testing elements on the module level, as well as testing conjunctions of modules and the final overall artifact. Since only the final artifact is safe enough for operation, the conjunction between artifact and procedures is only tested in the final stage of the project. As respondents acknowledged, the highest complexity could be found at this part of the process. In previous tunnel projects, given the momentum already built up by the technical artifact, actors had to choose between costly technical changes or locally adapted procedures that endangered interoperability.

Adding to the complexity for Delft was the fact that testing the entire tunnel in its final stages could only be done in a few days rather than the usual 6 months due to spatial constraints on placing temporary infrastructure. Foreseeing potential problems, the project team instigated an additional commissioning team that encompassed both members of the project as well as members of organizational entities that would eventually use the tunnel. It mainly sought to update final users on the progression of the project. It was this commissioning team that decided to conduct a so-called integrated procedural acceptance test (IPAT).

In the spring of 2013, the team conducted several days of scenario testing where representatives from all relevant stakeholders were asked to play realistic scenarios on a scaled-down prototype version of the railway tunnel. Game players were mostly members of the commissioning team with a few additional operators invited ad hoc. The prototype version included all real life software of the tunnel as well as realistic user interfaces that would be installed in control centers. Scenarios involved for instance the managing of train traffic and tunnel operations in case of a fire at the station platform. The intention of the IPAT was to show where the technical artifact and the procedures to use it would not match.

Table 6.2. Spoorzone Delft's innovation process

Spoorzone Delft			
	Product	Social	Institutional
Predevelopment	Predevelopment since 1990s but mostly on financing the project, rather than technical design of the tunnel itself.	Arena of involved actors is diverse and mostly external to the future members of the project. Inclusion is high as citizen involvement is sought and even instigated on their own. High amount of different perspectives (from local to national and on cost versus benefits).	Mostly a political arena, financial and political rationales play a larger role.
Take off	Design focuses on creating an overall grand design from which modules are derived. Structural complexity is at its highest and many disciplines are involved. Knowledge is readily available due to the involvement of contracted parties. Impacts of the desire to extend the project to two tunnel tubes instead of one on the complexity of the project are minimal.	Inclusion decreases as tunnel design becomes more technical and moves away from the more controversial issues such as the station and station area design. Number of perspectives still high as design involves many different aspects (safety, civil engineering, electrical engineering, etc.)	Special project team formed to bring together parties responsible for final commissioning. Mature institutional setting for dealing with this type of project with regard to the substantial design of the tunnel itself.
Acceleration	Structural complexity decreases as design takes place in modules. Design activity becomes more monodisciplinary and encompasses mostly already available technology. Knowledge availability is increased by extensive testing regime.	Social space becomes also more modularized. Early introduction of other late entrants due to the special commissioning team means that inclusion is relatively high compared to similar projects. However, low impact on the P-space.	System engineering practices mandate the institutional structure of the collaboration. Grand design restricts and controls the degrees of freedom of individual design teams.
Stabilization	Limited number of misfits creates additional complexity but project team is able to cope with it. Limited increase in amount of disciplines later on as tunnel is commissioned.	No significant changes in the S-space. For usual projects this is different as during commissioning new outside parties such as end-users and local and national authorities and inspection agencies.	Institutional setting becomes more formal as members of the commissioning team fall back to their own formal responsibilities.

Until the final commissioning of the tunnel in the beginning of 2015 the dynamics in the project are mostly impacted by those misfits between modules that previous tests could not foresee. The amount of new entrants to the process is less compared to similar tunnel projects due to the installment of the commissioning team in 2011. Also, the tunnel project is relatively shielded from dynamics elsewhere in the entire project. For instance, budget overruns and financial shortages after the economic crisis hits The Netherlands in 2008 mostly impact other areas such as the building of city offices on top of the station area and the building of houses around the immediate area but do not impact the design of the tunnel itself. An overview of the interview data is presented in Table 6.2.

6.2.3 DSSU

Before the start of the design phase of the DSSU project, the ideas behind the overhaul of the central node of Dutch network have been concocted in one specific department of ProRail. Although sector-wide there was an awareness of the potential lessons to be learned from the way the Japanese railway system was designed, detailed knowledge on the exact setup of the Japanese system was mainly located at the transport planning department. In retrospect members of this department can be seen as the driving force behind implementing 'Japanese principles' at Utrecht central station. Given the changes in the institutional structure before the start of the project (the sector had been split up in 2003, similar to the UK case), compartmentalization was high. Even within the organization of ProRail, departments more and more became separated entities. In this stage therefore little inclusion could be found and the project (although one could not speak of a project yet) was mostly mono-disciplinary. Dynamics were therefore absent in all three spaces. The vision for what Utrecht station should look like only slightly changed during this phase.

The project started when Utrecht central station had to be overhauled in light of expected increases in public and freight transport over the entire Dutch network. Appointed as a key project for the future Dutch railway network, Utrecht central station was to be expanded with an 8^{th} platform as well as see a entirely new station hall and station area. At the same time the sector was facing decreasing government spending and increasing spatial constraints on building additional infrastructure. This provided the transport planning department the opportunity to implement the 'Japanese principles' into the design of Utrecht. Since this department is responsible for collecting user demands (capacity requests by TOCs and maintenance departments) and translating these into a coherent set of functional requirements for the design teams, they were able to heavily steer the requirements to those implicitly demanding the use of the Japanese principles. In a context where financial resources were plenty, the actors within the sector always found ways to satisfy everyone's' demands when new infrastructural investments were planned or when innovations were implemented. In practice, this meant that every requirement, regardless what it entailed or who demanded it, became part of the project, as long as the Dutch government was willing to pay. Decreasing financial support and the overly complex system that resulted from this mechanism were reasons for ProRail, and especially the transport planning department, to start looking for ways to transform the system from an expensive and complex one to a more economically viable and simple one. This however entailed more constraints on the design of the infrastructure as trade-offs had to be inherently made. Surprisingly we see at this stage little dynamics in the P, S and I spaces. The institutional structure that is used for designing Utrecht is a standard one and to a large extent the initial involved actors are shielded from external demands. Also we see that what initially was deemed 'Japan' by the sector became more and more focused on solely infrastructural measures. Hence, structural complexity of the innovation artifact remains low due to the abstraction of the theory by which the department talks about the innovation. Also, few *disciplines (few perspectives)* are involved in the first stages of the design process of Utrecht.

The transport planning department is responsible for collecting the user requirements for the new station layout and translating these to a coherent set of demands to the project team responsible for engineering and designing the solution. In these stages of the process we see that the project team for DSSU and the engineering companies they contract are shielded from the rest of the organization. At this moment, the rest of the organization has little insight in the far-reaching impacts of the decisions the project team is taking. Given that the project team is using standard procedures for designing railway station layouts but are pushed by the transport planning department to use Japanese principles, they have to work around the many constraints that this places on their design space. The project team more and more incorporates projects occurring elsewhere in the network to adhere to the demands of the transport planning department as well as communicate design decisions that do not adhere to prescribed protocols to the safety departments of ProRail and the Dutch railways. At this stage there is little feedback from these departments, which initially keeps the complexity of the project low. For instance, the project team assumes that freight trains will be diverted from Utrecht over a newly upgraded corridor in the east of the country. The extent to which this upgrade is indeed being built is not taken into account later on.

At the same time, winter disruptions during the years the project team is involved in designing DSSU cause the sector to try to cherry-pick elements from the innovation to solve pressing current issues. Other projects seek to remove railway switches earlier on, an integral part of the Japanese principles, to counter the problems at Utrecht with switch failure during wintery weather. Whereas for the innovation itself removing switches has a completely different purpose, in isolation the measure can also be used for other purposes.

Whereas the initial process of DSSU was quite stable with respect to dynamics in the product, social and institutional space, in the final stage we see that dynamics increase heavily. The safety departments are overloaded with deviations from design protocols and all of sudden realize that they cannot simply assess these, as they would usually do. Furthermore, the rest of the organization realizes that the project had been implementing measures deemed not yet implementable. For instance, the project team uses signal optimization to adhere to the demands of the transport planning department but this measure is being researched at the same time for its consequences on safety. Only when the final designs are communicated to the rest of the organization do they become aware of the impact of implementing Japanese principles.

At the end of 2012 ProRail organizes three workshops to deal with the sudden controversies surrounding DSSU, a few months before the actual building commences. The workshops show that many dilemmas have appeared that previously were neglected by the directly involved stakeholders and that the participants of the workshop find it hard to find the right institutions to deal with these dilemmas. Here we see that the design of Utrecht still changes, from a highly ambitious move towards a Japanese inspired station (initially removing 90% of the switches) to more moderate change. Given these changes, the project is not able to provide all the initially demanded functions properly. For example, at the beginning of the process the transport planning demanded an infrastructure that would allow trains to follow each other with a maximum of 2 minutes headway. Given the influences of new entrants demanding changes to the design, this ambition had to be lowered. Table 6.3 shows the findings per phase for the product, social and institutional space of the process.

Table 6.3: DSSU innovation process

DSSU			
	Product	Social	Institutional
Predevelopment	Mainly a transport planning project, looking at the separation of corridors to avoid secondary delays and to 'simplify' the system. Very abstract theory of the artifact and few languages and disciplines involved. More in depth study of Japan where the department deemed a simple and robust system can be found. Gradually opening of product space: signaling, timetabling, maintenance, traffic control all become part of the product description.	Initially, actor network consisted of one department within organization (transport and timetabling). Due to abstract theory of the artifact, they stays under the radar. Inclusion is low as well as the amount of perspectives. Capabilities needed mainly focus on studying the concept in more detail.	Split up of railway sector into infrastructure manager and train operating companies; more clear division in incentives. Less coordination between actors both internally and externally. Individual departments better able to singlehandedly search for promising avenues of improvement.

Take-off	Large funding for rebuilding Utrecht Central Station. Mainly an infrastructural investment but also serves as opportunity to implement 'Japanese principles'. Product space becomes partly technical and partly sociotechnical, but complexity remains low compared to later stages. Focus is solely on infrastructure components of the innovation. Other components of 'Japan' are neglected.	Organizationally stipulated opening up of S-space as transport planning department has to collect user requirements for the project. However, transport planning prioritizes these requirements to meet their own demands, this is new to the sector (and unknown to them)	Institutional setting used for building Utrecht Central station is standard to the sector. Transport planning department is in the lead for determining the set of requirements of the final design of the station layout.
Acceleration	Many constraints on the design space, potentially increasing the structural complexity. However, few disciplines involved during design process. Project team can deal with the constraints by expanding the design space in other areas (deviating from protocols and taking into account other projects).	Project team is designing DSSU. Inclusion is low since few other parties are involved. Little variety in perspectives, solely focused on achieving prescribed set of requirements.	Highly formalized institutional setting for designing station layout. Communication from project to rest of the organization mostly indirect through drawings and documents. Communication on deviations from protocols and taking into account other projects is mostly unidirectional (from project to the rest of the organization)
Stabilization	Implementation is more chaotic. Many second order effects of the niche on the regime are found but are qualitatively different and need to be properly studied. This happens while design, build, and testing processes are running simultaneously. Little buildup of momentum: design changes are becoming more frequent. Additional pressures on project team during design process due to large disruptions around Utrecht, exaptation: principles serve new functions. Removing switches is not done to separate corridors but to avoid winter disruptions (cherry picking): regime actors try to introduce parts of the innovation to meet other pressing demands.	Social space 'explodes'. Many departments enter the social space, especially those from the standing organization and those from the innovation focusing on neglected functions (safety, traffic control). Many different perspectives need to be aligned which needs new but lacking capabilities	Design rules become more prominent as detailed design is taking pace but remain stable and have second order effects. Many design changes have to be sanctioned by the safety echelons of all organizations. They are overloaded. Other projects, on which the DSSU team relied for maintaining low complexity, see dynamics that impact DSSU. Assumptions used during design process no longer hold. No clear rules for dealing with these new phenomena.

6.3 Analysis

If we compare the three cases we see distinct patterns in the dynamics of the P, S and I spaces over time as well as how dynamics in each spaces impact other spaces. In general, the timing of volatility in these spaces seem to be mirror images for the TMS and the DSSU project, with the SZD project somewhere in between. In this paragraph we first do a cross-case comparison per single space of the PSI framework, looking at how for each case the dynamics in a single space was different. We synthesize these findings into an overall analysis of the different macro-level patterns found where we also take into account the mutual impact different spaces can have during a single innovation process.

6.3.1 Institutional space

DSSU is distinct in that the project was a combination of both a traditional overhaul and extension of a railway system and the introduction of Japanese principles for redesigning the infrastructure. In addition to more complexity in the P-space, to which we will turn later, this also introduced a less well-defined institutional structure for managing the interactions between involved stakeholders. In essence the institutional structure used in this project, which was compounded with highly new design principles, was one used for traditional projects. This was also the case for the TMS project and the SZD project but here the institutional structure was set up around the project involving solely these innovations. For the latter two cases the institutional structure, such as project guidelines and rules on who to involve, were completely adapted to bring about the innovation. In the TMS case we see some changes to the institutional structure after take-off, such as a more compartmentalized stakeholder arena, but this was in congruence with shifts in the P-space where the innovation became structurally decoupled from other innovations. Crucial here is the difference in how the take-off of the three innovations took shape. Whereas the windows of opportunity in the TMS and SZD cases stipulated the implementation of solely their respective innovations, for DSSU the innovation had to be latched on to an other ongoing project, which was the renovation of Utrecht Central station. We therefore see that whereas two projects become intertwined (implementing Japanese principles and rebuilding Utrecht) the standard institutional structure remains the one applicable to the latter.

In addition, for the DSSU case ownership and agency is more diffuse. At the two other projects, a clear project ownership becomes visible after some dynamics in the earlier stages of the process. At TMS the innovation is put in a concrete project, separate from other innovation projects. And even when the project is handed over to other owners, such as during the transition to the infrastructure delivery department, agency is still in the hands of the same people. Also we see that in both TMS and SZD the ties between stakeholders is much stronger earlier on, purposefully instigated by the initial actors, to allow for more overlap and communication. Examples are the introduction of end-users in both cases and the invitation for unions, municipalities, emergency services to participate. At DSSU we see that project ownership switches from stakeholder to stakeholder over time and far less overlap exists in how the stakeholders are related to each other.

What we finally see regarding how institutions shape over time is that whereas political dynamics largely played a role in the less systemic innovation processes, at DSSU there were little dynamics earlier on. For the project owners in the beginning the way these Japanese principles would have to be implemented was quite straightforward: they would specify the functional requirements of the new station infrastructure so as to demand the subsequent design team to use these principles. In addition, it was assumed that this could be done using the traditional way of organizing overhauls of railway stations. Later on, when dilemmas appear that the broader organization was not aware of before, these traditional institutions were seen as ill fitted to deal with them. Many respondents during the workshops of DSSU noticed that usual ways of managing the project, such as the way decisions would be made or the way design protocols would bound the design team, could not help in diminishing the problems that the organization encountered in later phases. Most notably, the ad-hoc instigated workshops were a corroboration of this statement: right before commencing the actual building of the station, stakeholders from all involved departments both within and outside ProRail joined to together to freely discuss issues and institutions to deal with these issues.

6.3.2 Social space

DSSU is also notably different for the way the innovation itself had to be financed. Given that it was merely design principles, however systemically different they were, they needed little to no financial resources. The main discussion in earlier phases involved overhauling Utrecht or not and with the Japanese principles it would mean, theoretically, that the overhaul would become much cheaper. The TMS and SZD projects involved costly innovations in itself and hence the dynamics were much more political and public in nature. As innovations they had to compete with other solutions for the needed resources and therefore they had the involvement of a range of stakeholders earlier on, even before the actual design of the innovation started. However, given that to some extent the innovation were already there (at TMS the process started with an innovation and not a problem), the making of a business case might be deemed easier than is the case for DSSU.

Also we see that inclusion is far less at DSSU in earlier stages. Distinct from the other two projects the initial driving force behind the innovation is a single department that remains relatively a niche environment. At that moment there is little opportunity for other stakeholders to influence the process. Given that upfront investments were needed, the initial driving forces for TMS and SZD had to open up the S-space to allow other stakeholders with relevant resources to be included. In addition, given that the innovations were to become highly inert later on, the phases during and after take-off saw an influx of stakeholders otherwise involved only later on in the process. This led to many perspectives and languages to be involved in the innovation process. For example, the introduction of emergency services and end-users at the SZD project meant that the functional requirements could be much more explicitly communicated during the design of the technical specifications of the tunnel. These requirements were highly varied. On the other hand, we see little to no changes in the constellation of actors and their perspectives and languages later on. So for TMS and SZD the S-space slowly settles over time, albeit that for the SZD project the inherent overlap of the tunnel project with the regime system it is built in causes some additional stakeholders to enter the arena later on, such as notifying bodies and other inspectorates that need to officially commission some aspects of the tunnel.

At DSSU we see the opposite: from a single department, to a more broader set of actors who design and engineer the innovation and the overhaul of the station to finally a full range of qualitatively different stakeholders at the very end. Whereas perspectives in the beginning were mainly focused on infrastructural changes and the realization of smaller headway times between trains, later on perspectives on safety, flexibility, operability and costs enter the S-space. A noteworthy phenomenon we saw at DSSU is also the entrance of stakeholders that are not so much interested in adopting the innovation elsewhere (as was the case for TMS) but much more interested in picking elements from the innovation and use it for qualitatively different problems. So in both cases did we see new entrants that could benefit from the innovation but for DSSU the benefits stemmed from using aspect systems of the innovation and not a scaled-down subsystem of the innovation.

6.3.3 Product space

From the outset we have said that the structural makeup was different for the three projects, but only in retrospect. TMS shows an interesting feature that DSSU did not have. Whereas the initial ideas revolved around installing a multitude of innovations somewhat concurrently, ambitions became much more modest over time. Also, to ease the implementation Network Rail chose to buy the system off-the-shelve. Here we see that the internal architecture of the innovation more and more becomes a collection of highly interdependent elements that no longer can be easily changed at will. For this reason we saw the upfront involvement of stakeholders since only then crucial changes could be made. For the SZD we saw the system engineering principles applied to deal with these issues, causing more modularization towards the middle of the process. What this shows is that the innovation itself over time gains momentum and changes become more costly and less feasible over time. For DSSU this did not seem to hold to the same extent. Later on, infrastructure designs were changed due to pressures from new stakeholders, other projects sought to cherry-pick elements of it to solve other issues and many new dilemmas about the innovation became apparent. Hence, DSSU seems to be much more dynamic in what constitutes the innovation then the other projects, and this dynamic is mostly found at the end of the process. During predevelopment phases and take-off the Japanese principles are a coherent set of elements, albeit more mono-disciplinary than what it turned out to be. As later on the S-space opens up we see that this has more repercussions on the dynamics in the P-space since the innovation was not able to built up enough momentum. As design principles entailed a set of measures that could just as well be implemented separately, there were more opportunities to manipulate the innovation along the way.

6.3.4 Cross-case comparison of P, S, and I

Studying the dynamics in the P, S, and I spaces in three different cases we therefore see striking differences over time. However, although the cases involved different types of innovations or innovative artifacts, we do also see some resemblances. Firstly, all three cases involved a clearly demarcated project during all or most of all stages of the process. We saw that initially all three cases the implementation of the innovation was done using existing standard procedures for bringing about such projects. At Spoorzone Delft they used standard system engineering principles for structuring the design and implementation process, similar to the approach used by Network Rail. Although less structured at the DSSU case, still initially they used approaches standard to renovating and overhauling infrastructural layouts around railway stations. Furthermore all three cases needed a window-of-opportunity to go from predevelopment to take-off. Tipping points were for instance the sudden pressure on Network Rail through the McNulty report and the beginning of CP5, the nationwide ambition for increased capacity and the need for doubling the tracks at Delft, and the long awaited upgrade of Utrecht central station to allow for the implementation of Japanese Principles.

Although all three cases needed a window-of-opportunity, these windows are however qualitatively different. DSSU stands apart here as the initial innovation had to be combined with the usual overhaul of Utrecht that was already planned. In that sense, the Utrecht project became a combination of a standard project about renovating the station and its surrounding area and the implementation of radically new design principles from Japan. In the other cases the project team had a simpler task as they could solely focus on the innovation itself.

Most importantly however is the evolution of the innovation artifact over time, which seems to have a pervasive interaction with the dynamics in the social and institutional spaces as well. Especially the Network Rail and DSSU case seem to be mirror images to this respect. Whereas rapid changes in the artifact, actor networks and institutions were happening at the beginning of the process at Network Rail and few dynamics can be found in the stabilization phase, for DSSU this was entirely different. In retrospect we saw that the three innovations were structurally different and hence the reason for their inclusion into this multiple case study. Looking from the start of their respective project we see that sometimes these differences are not exogenous but instigated by the project members themselves. If we compare DSSU with the Network Rail case this becomes visible:

Whereas at DSSU the team engineered the solution themselves, at Network Rail they bought the innovation off-the-shelve. For the involved stakeholders at DSSU this meant that early on in the process the innovation was highly malleable. Given that the design was their own responsibility they had more freedom in changing the innovation compared to Network Rail. Buying it off-the-shelve means that many structural lock-ins in the innovation are already present and the project cannot freely change the product they procure. The respondents at Network Rail found that involvement of end-users during testing highlighted this tension between user-centered design and buying off-the-shelve. However, by buying off-the-shelve the project immediately had to find resources in the form of financial and political support. This increased the visibility of the project early on and created many dynamics in these stages: solutions, solution-owners and institutions were discussed and changed. In comparison, at DSSU in the early stages the involved stakeholders could stay relatively under the radar since their innovation involved a mere theory on how to design existing components of the railway system. Hence, they needed less financial investments but merely had to await a window-of-opportunity that needed this new ‘theory’. We see that in the DSSU case the innovation builds up far less momentum over time and is subject to changes even in later stages of the process. A striking difference here is also the way quick wins are realized. At Network Rail operating centers can choose for early deployment by having some of the TMS functionalities already installed, functionalities that the final TMS will also provide. At DSSU, the early removal of switches that was discussed in 2012 is part of the Japanese principles but in isolation serves qualitatively different goals than it would combined with signal optimization and corridor separation.

On the other hand Network Rail and DSSU differ in the amount of overlap their respective innovations had with the existing system and with concurrent innovations. In essence, the traffic management system in the UK was laid over the existing network through the use of newly designed interfaces. Also, they decoupled the innovation from other promising innovation that were initially deemed crucial to implement. A Driver-advisory system (DAS), which could provide synergy in combination with the TMS, and the installment of ERTMS have been purposefully decoupled from the TMS project. This created the effect that decisions on the TMS had no impact on other projects, and more importantly: vice versa. So as the project moved away from a predevelopment phase towards acceleration and more and more interfered with ongoing operations in the current system, the technical makeup of the innovation was only slightly impacted by external pressures.

For DSSU we see a totally different picture. The elements of the initial innovation (signal optimization, switch removal, separation of corridors) are also part of other ongoing innovations or part of the current organization. Whereas the project intends to use signal optimization to increase capacity and robustness, the current organization only uses it for safety reasons. Hence, in later stages the dilemmas appear between using signaling to optimize capacity versus safety. Also we see that DSSU more and more becomes dependent on other projects, completely contrary to what happened at Network Rail. This creates the effect that volatility in P, S and I space occurs in later stages instead of earlier stages, at the moment when the project more and more interferes with ongoing operations.

The SZD case in this analysis forms a middle ground between the two extremes of the Network Rail and DSSU case. Dynamics were profound in the beginning since the tunnel inherently needed much upfront financial and political support but the dynamics were less due to the lower risks involved. In essence, building a tunnel is not really an innovation and hence less uncertainty can be found when deciding on building one. We also see that dynamics can be found at later stages when independently designed modules are brought back together and when the system interacts with the current system. However, these dynamics are not as extreme as for the DSSU case. We see that a tunnel is far less malleable in later stages than the innovation of DSSU. The SZD case is also an interesting one since many of the volatility in later stages had been front-loaded by installing a special commissioning team earlier on and by having a more rigid testing regime than usual. These notions are interesting for Chapter 8 on the use of gaming methods.

What is crucial to understand after analyzing the three cases and serves as impetus for the next chapter is that the innovation process at DSSU seemed to be much more open and that this openness increased over time. Whereas at TMS and SZD the initial dynamics could be attenuated by putting up boundaries on what could be designed as well as who could be involved, DSSU was much more fluid and ambiguous as a process. Although DSSU itself was a project indeed, if we look solely at the trajectory the innovation took from an idea into being installed during the DSSU project, the actors that were involved and the impacts that had on the technical design of the innovation were not merely stemming from the stakeholder arena internal to the project. Hence an analysis of the dynamics of the innovation processes needs to take into account that the project-level is not the right unit of analysis.

6.4 Synthesis

If we both look at the differences per space between the three cases as well as how per case the different spaces impacted each other differently, we can make a set of five claims that might be valid beyond the three cases underlying the research of this chapter as well as provide input for both the single case study of the subsequent chapter and the chapter on gaming simulation in Chapter 8. These five claims involve the unique macro-level pattern of a systemic innovation process and how its systemic nature created this macro-level pattern:

1. The key factor that allowed us to distinguish three different innovation processes was the location in time of volatility in the P, S and I spaces. DSSU saw this volatility mainly in later stages of the process, with much stability in the beginning, while TMS saw the exact opposite

2. The location in time of volatility in the P, S and I spaces is dependent on the ability of the innovation to build up momentum. Low momentum innovations such as DSSU are more prone to changes during the implementation process. High momentum innovations tend to mirror to a lesser extent the dynamics in the social space, i.e. new entrants to the social space cannot freely impact the P space.

3. The internal architecture of the innovation influences its ability to build up momentum. Innovations where the elements are highly interdependent implicate a quick buildup of social and institutional arenas to provide the innovation with resources. Low epistatic innovations (were interdependence is less) are able to remain an abstract 'idea' for a long time and also are more prone to cherry picking by the current organization to solve other pressing issues.

4. Low epistatic innovations such as DSSU are often collections of many qualitatively different elements. The design responsibility of these elements is often spread over different actors with different languages, incentives and decision rules. This increases the projects' responsiveness to outside pressures and the volatility of the implementation process.

5. Momentum defies outside pressures. The strength of these pressures is however determined by the amount of shared elements in innovation and the current system. Little overlap causes, ceteris paribus, innovations to invade the regime system more smoothly. More overlap causes innovation elements to be influenced by both the directly involved actors themselves as well as by the regime system and concurrent innovations. This creates many second order effects at the end of an innovation process when the innovation and regime and other innovations start to more strongly interact.

Given the aforementioned claims on how the design of an innovation impacts the process of implementing it, we can provide a typology of different innovation patterns in Table 6.4.

Table 6.4: typology of different niche-regime configurations

		Innovation-regime overlap	
		Little overlap	Much overlap
Internal architecture	Tightly coupled	Block add-on (Network Rail)	Hybrid (SZD project)
	Loosely coupled	?	Shotgun-invasion (ProRail)

We set apart the TMS and DSSU project according to the internal systemness of the technological artifacts' components and name them either blocks (tightly coupled) or use the analogy of the spray of a shotgun (loosely coupled). Also the way the innovation artifact is to interact with the regime system can be used to distinguish the two projects: we term these either add-ons (when there is little overlap and the artifact merely attaches to the regime system) or invasions (where innovation elements replace regime elements).

A highly abstract graphical representation of the volatility in product, social and institutional spaces for both a more radical innovation and a systemic innovation is represented in Figure 6.1. This figure synthesizes all the findings of the multiple case study of this chapter into a graphical comparison. Note that time and volatility are relative terms and are merely used to show the qualitative differences between the two processes, not the quantitative differences. For instance, one of the processes might inherently take longer periods of time or might always see higher volatility than the other process. Since we had no analytical framework to compare the level of volatility, but only its location in time, our final results neglect these differences between the TMS and DSSU projects.

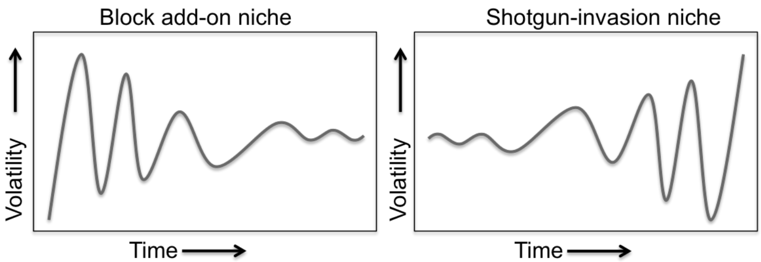


Figure 6.1: niche accumulation volatility over time for two different niche innovations

6.5 Conclusion

In comparing different innovation processes we show that systemic innovation processes are qualitatively different in that the volatility is located at the back end of the process. Whereas Network Rail in the predevelopment phase decided to keep the solution internally tightly coupled (by buying off-the-shelve) and externally loosely coupled (by purposefully creating independence of other concurrent niches and regime elements), The DSSU project strived towards an innovation that was internally loosely coupled and externally tightly coupled, seeking synergistic effects on many fronts. This caused the solution to respond rapidly to outside pressures, placing much more pressure on the management of the social space. In the end, we feel that two innovation process patterns became visible. Firstly, to create highly epistatic niche innovations (where the elements of the innovation are highly interdependent for the overall functioning), many parameters have to be decided upon in the front-end of the innovation process. For less epistatic niche innovations, the specifics of the P-space only start to play a role in the back-end of the innovation process as relative loosely coupled elements start to interact with highly interlinked regime elements. Take for instance the element 'railway switches' in the DSSU case, which were both important for the innovation and for the regime system. While for the innovation the railway switches had to be changed in order to increase the robustness of the network, for parts of the regime system these switches were important to maintain flexibility and for others to cut maintenance costs. For this reason, rather than a chaotic start, for these types of innovations we propose that the chaos builds up rather than breaks down as an innovation is implemented in the regime system and more innovation elements become the focal point of the regime system as well.

The findings of this chapter show the pervasive role of the technological makeup of an innovation on the process of implementing it. In large part, volatility of the process and its location in space and time was a direct result of the extent to which an innovation became a fully specified product with a fixed design early on (TMS) or stayed an abstract idea with many elements that were loosely coupled (DSSU). Current literature on systemic innovation processes, especially the MLP focus on the dynamics of the process without providing scholars any framework to incorporate the role of the innovation itself. In many studies using the MLP, innovation scholars do not discuss the way different innovation elements are linked to constitute the innovation artifact, how this innovation can change over time, and how this change impacts the innovation process and vice versa.

For instance, if we relate our findings to the MLP literature we see that the way niche innovations invade or add-on to the regime depends highly on the structural and technological makeup of the innovation that is concocted in the initial niche environment. Block add-on innovations invade a regime with a different pattern than Shotgun-invasion type innovations. Block add-on niches adhere to normal dynamics found in innovation processes (Cheng and Van de Ven, 1996) where highly chaotic front-end processes slowly settle for more stable implementation processes. Shotgun-invasion niches (resulting from their systemic yet loosely coupled properties) show a reverse order of chaos and stability (see Figure 6.2 and 6.3).

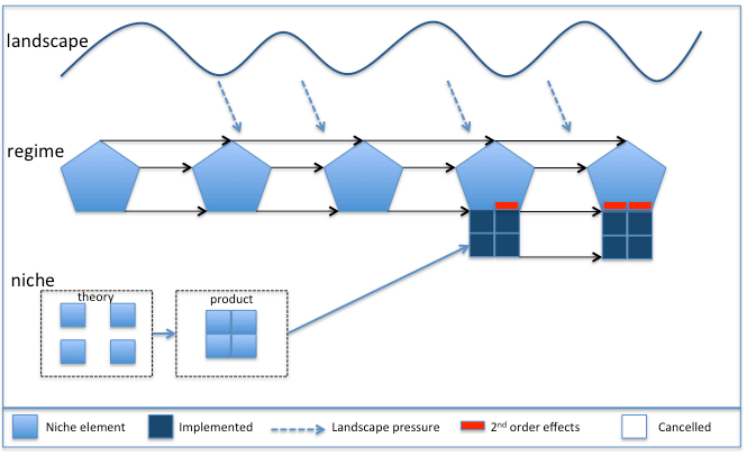


Figure 6.2. Niche accumulation of 'block add-on' niche

Block add-ons typically build up momentum early on in the process, where a mere theory on how it should function and what it should look like is rapidly translated into a concrete artifact. Given its high momentum and little possibilities for the regime to cherry-pick elements from it, the interactions between regime and niche mainly are one directional: adding-on the new technology stipulates 2nd order effects in the regime and not vice versa. At later stages of the process the innovation artifact itself barely changes.

Shotgun-invasion niches are different in that the innovation remains a theory for longer periods of time and builds up less momentum. It is also easier for the regime to cherry-pick elements of it to meet pressing demands from the landscape. Now, the elements that are introduced to the regime, i.e. implemented, need to survive in isolation and this changes the makeup of the innovation that is subsequently deemed implemented.

The two different macro-level patterns this chapter uncovered implicate two different ways of handling systemic innovations in networked infrastructures. Crucial here is the extent to which an innovation autonomously creates momentum. For internal tightly coupled innovation, momentum seems to grow as money, time and knowledge is invested.

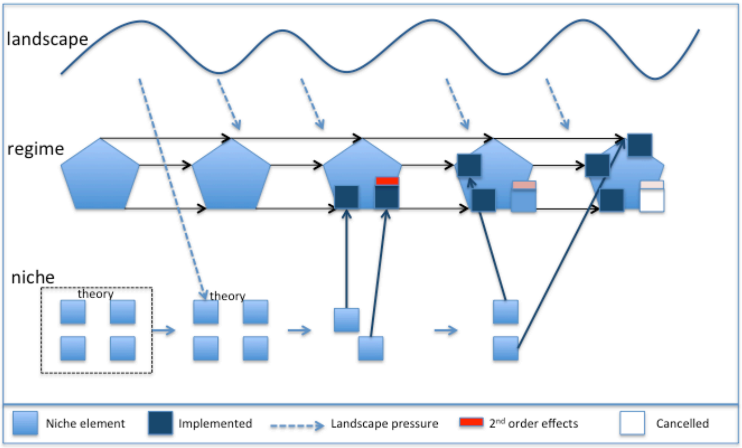


Figure 6.3. Niche accumulation of 'Shotgun-invasion' niche

Because of this momentum, the volatility of the implementation process dampens over time. For less tightly coupled innovations, the implementation process is far more volatile and the dynamics around this process increase the more the process is in the, not so aptly named, 'stabilization' phase. In properly handling these innovation projects, the latter far more demand from so-called 'change inclined regime players' (Rotmans and Loorbach, 2009) to front-load the complexity inherent to these innovation processes. As these processes are started with abstract theories of the artifact, innovation stakeholders should ideally early on test these theories in protected environments to uncover the second order effects of innovation-regime interactions that would otherwise only appear in the final stages of the implementation process.

However, most importantly we have shown what makes systemic innovations of the kind we are interested in (loosely coupled measures that intend to change the makeup of an existing system) so unique and problematic. We propose that influencing the volatility of the process could be an interesting functionality of gaming simulation. We see that in the DSSU case the involved actors at later stages were no longer able to deal effectively with the dilemmas they were confronted with. Gaming simulation could then be designed as such to front-load this volatility, for instance by introducing stakeholders that would otherwise only enter the S-space in later stages.

The limitations of the analysis we used however force us to use the results with some caution. Firstly, we used different methods for different cases for practical reasons. This might distort our view of the true dynamics over time in each process. We still feel that the conceptual level of our claims, looking for macro-level patterns, is less impacted by this shortcoming. Also the processes we studied were located in different contexts, and even in different institutional settings and countries, which might endanger the validity of our claims. Our analysis focused solely on the project-level and did not account for contextual influences. So for instance, it might be that the context in which TMS was implemented prohibited the entry of influential stakeholders later on, as was the case for DSSU, making the TMS case potentially just as volatile at the end of the process. This does not however lead us to completely discard all our propositions since our analysis was on the project-level (the same for each case) and how the project could cope with contextual pressures over time. Thus indeed the TMS case has not been tested to the same extent as the DSSU case, as it did not see much contextual pressures later on. Still we fairly assume that given the momentum the artifact built up and the way this impacted the influence of contextual pressures mid-stage of the process, this will also dampen contextual influences at later stages. We add here that we were interested in qualitative differences between projects and not so much in quantitative differences. A similar project as TMS might see volatility at later stages but this would be lower in scale had the similar project involved a DSSU-like innovation.

In addition, we need to be wary of the normative claims we base on our analysis. Influencing volatility is potentially problematic. Firstly, it is a very abstract notion and provides little guidance to the gaming simulation professional. There are potentially different mechanisms at work that create this volatility and we have yet to figure out what these mechanisms are. Secondly, there might very well be a method to this ‘madness’. In the end DSSU got implemented and the fact that volatility was located at the end might be intentional or at least to some extent functional. Looking at the underlying mechanism might reveal the functional aspects of the fuzzy back-end of innovation processes. For this reason, Chapter 7 will delve deeper in the DSSU case and uncover the underlying mechanisms that explain the macro-level pattern found in this chapter

7 Uncovering the Driving Micro Mechanisms at Play During Systemic Innovation Processes: the case of 'Doorstroomstation Utrecht'

In the introduction of this thesis we stated our intention to both construct a global model and a local model of a systemic innovation process. This intention stems from Van de Ven and Poole's (1989) description of innovation processes as being describable on the pattern and the mechanism level. This description again emanates from a broader ontological and epistemological trend in innovation research where complexity, temporality and emergence play a large role. This means that, just like for instance the flocking of birds, these processes can only be understood if one studies the emergent patterns of a collection of actions well as the concrete rationale for the individual action itself. The previous chapter helped in mapping the emergent pattern, which in this case involved a pattern of increasing volatility over time. As we concluded there, this pattern is distinct from more traditional innovation processes, whether they are incremental or even more radical. While the previous chapter showed that systemic innovation processes are different, there is still little knowledge regarding why these processes are different. We proposed that the technological makeup of the innovation artifact played a large role, as the so-called epistasis of the innovation (or the ease by which one can add or subtract an element from the innovation without immediately impacting the functionality of the innovation) determined whether or not an innovation process would build up momentum over time to resist regime pressures.

What the previous chapter lacked and what this chapter intends to add to the analysis is a more in-depth understanding of why a systemic innovation process is different from other innovation processes. We provide this depth by focusing on three different phenomena that the previous chapter left untouched. Firstly, in this chapter we introduce agency to the analysis. Innovation processes are in the end the result of individual actions of actors that want to change or improve something. Their intentions, incentives, decision making processes and resources hence play a role in driving forward or blocking an innovation process. We saw for instance that the technological makeup of the innovation and its dependency on other innovations differed between the DSSU and the TMS case. Although partly a difference grounded in emergent patterns beyond the control of the actors involved, the difference was also the result of intentional actions by these actors. During the TMS case, there were points where a collection of actors decided to build interfaces between the innovation and the regime system. At the DSSU case the involved actors could have decided to implement the innovation as one 'block', in a take-it-or-leave-it kind of deal, similar to the way the traffic management system was implemented at Network Rail. However, for the DSSU case they did not. This chapter will do justice to these immediate micro-level actions more fully, as agency becomes a part of the analysis.

In the methodological chapter we additionally stated that more depth not merely involves looking in more detail to the process, i.e. zooming in, but in a sense also broadening the scope of the analysis. By taking into account the history and context, making it a more intrinsic part of the analysis, we better understand why on a macro-level we found the patterns we found in Chapter 6. These two phenomena will help embedding the micro-level actions of actors in their historic and spatial context. Actors have a history, learn from other historical innovation processes, or compete with other innovation processes. In addition, institutions, formal and informal, are the echo of historical lessons and bound what individual actors can and cannot do.

The problem with depth is however that the analysis runs the risk of retrieving knowledge on and an understanding of the process that is so local as to become untranslatable to other similar processes. An innovation actor might have decided to wait for implementing the innovation because of a specific event occurring, a specific norm he or she applies, or a specific historical lesson the sector has learned a decade earlier. Then saying something in more abstract terms about this pivotal point in the process in order to understand or improve a process in a similar setting or in a similar network-type industry becomes highly problematic. Therefore, the level of analysis that this chapter focuses on is on the level of the mechanism. The goal of this chapter therefore is to find a set of mechanisms that are at work during a systemic innovation process. From the outset we term those mechanisms relevant if they adhere to three restricting factors: firstly, they need to help in explaining the progress of the innovation process in general. Secondly, the mechanisms need to help in explaining the general pattern of increasing volatility we found in Chapter 6. Finally, the mechanisms need to be a combination of immediate actions by actors and the pervasive impact of context and history.

The focus on mechanisms not only helps in better understanding systemic change in network-type industries such as the railways, mechanisms also function as a bridge between the results of Chapter 6 and the analysis in the next chapter, where we will look at how gaming simulation can contribute to innovation processes. We stipulated that gaming simulation might help in better tackling the problematic volatility increase we saw at the DSSU case. However, the link between the employment of a gaming simulation and the volatility in design, actor arena and institutions is too abstract without any knowledge of more micro-level mechanisms as a mediating factor. In addition, the analysis in this chapter also helps in better understanding the involved dilemmas in manipulating any volatility. As we stated in the conclusion of the previous chapter, there might be constructive forces at work that both resulted in the implementation of the DSSU project as well as the increase in volatility in later stages: the so-called 'method to the madness'. Results from this chapter therefore may warrant a more nuanced look at how gaming simulation may improve systemic innovation processes, or even may have negative impacts.

7.1 Method and case description

In this chapter we again look at the DSSU case but now broaden the scope in space and time. We look at how ideas inspired by Japan have over time gotten foothold at the redesign of Utrecht Central station, where the DSSU project is the concrete manifestation of this process. This results in an analysis that looks much further beyond the initial predevelopment phase of DSSU, which was around 2008-2010, but towards initial visits to Japan by entirely different organizational entities within the Dutch railway sector (around 1997). Also, we look at actors that were not directly involved in the DSSU project and how they behaved during these years.

Whereas in the multiple case studies from the previous chapter we had a predesigned analytical framework and could hence quickly focus on specific respondents for our interviews, this analysis used a snowball method to pursue interesting phenomena even beyond the initial scope of the analysis. The seed interviews were those resulting from the workshops in 2012 to which we provided support. As we mentioned before, these workshops were organized to tackle a multitude of dilemmas that emerges relatively late in the process. From this initial set of interviews we derived interesting phenomena and additional respondents who, according to the respondents, could provide additional insight to these phenomena. As Chapter 5 showed, the pool of respondents eventually covered both those involved directly in the DSSU project as well as respondents more distant to it but highly relevant for the analysis.

In getting from the interview data to a set of relevant mechanisms we used an approach where we iteratively induced mechanisms from the data, related this to the macro-level pattern we found in Chapter 6 and went back to the data from the interviews again. The raw data stemming from the interviews resulted in two classes of information. Firstly, interview respondents elaborated in concrete terms on a timeline of events that occurred over the course of the innovation process. This timeline of events was enriched with detailed data found in project documents. Secondly, respondents noted on certain recurring tendencies in the sector, specific problems, or ways they themselves acted during the process. This latter data finally helped in finding mechanisms. In the analysis underlying this chapter we then related these mechanisms to both the concrete timeline of events and the pivotal points during the succession of these events as well as the macro-level pattern of increasing volatility. In addition we continuously related the theory we built, consisting mainly of mechanisms, to already existing theoretical frameworks. This process of analysis was continued until no pivotal points were unexplained and the found mechanisms were deemed sufficient to explain why the DSSU case saw increases in volatility rather than decreases. In essence, the theory building resulted from a constant interaction between the raw data from the interviews and project documents, the theoretical frameworks from Chapter 3 and the macro-level pattern of Chapter 6. In Figure 7.1 we show how this theory-building occurred. This figure also shows how the theory-building was not a purely grounded theory approach as both theory and existing empirical findings structured the building of a new theory.

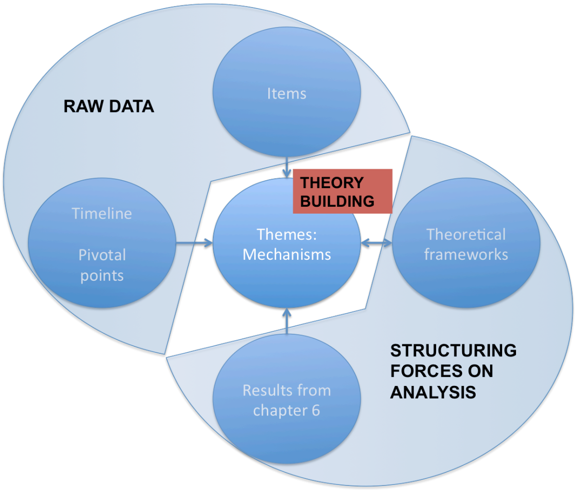


Figure 7.1. How theory-building resulted from the juxtaposition of raw data with structuring forces.

Although highly iterative, the report on this analysis in this chapter portrays a linear move from data to items to mechanisms to patterns. We firstly start by ordering the data we retrieved from the interviews and the documents to make a timeline of events. We then analyze this timeline to see remarkable events that seem to be pivotal in explaining the overall course of the process. Then we return to the raw interview data to retrieve information on how interviews respondents mentioned how they themselves sought to impact this process, for what reasons and with what result. Or, if they elaborated on more general tendencies within the sector, how these tendencies might help explain the specific timeline of events we found. Hence, after presenting the pivotal points in the timeline, we use a slightly adapted grounded theory approach to distill from the raw data items. These items we then try to fit into mechanisms, which we explain the middle part of this chapter. This chapter ends with two sections that embed the mechanisms we found in both the theoretical frameworks already explained in Chapter 4 and the pattern found empirically in Chapter 6. As said, in reality this embedding was done constantly during the analysis itself and thus embedding impacted the analysis and the results as well as the results impact the embedding. In Figure 7.2 we graphically depict the outline of this thesis.

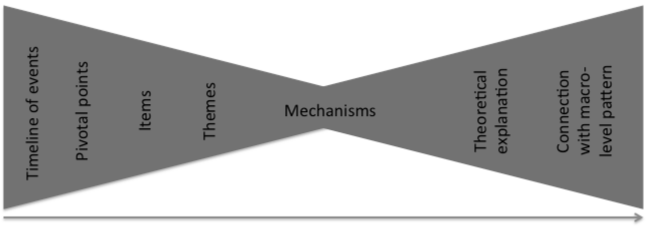


Figure 7.2. Outline of this chapter, from a timeline of events to relating mechanisms to the macro-level pattern

7.2 Data

The interview resulted in two classes of interesting data. Firstly, it helped in building a timeline of events. The building of the timeline was supported by organizational and project reports as well. Secondly, the interviews resulted in phenomena explaining why things happened as they did. In the first part we show the timeline of events, in the second part we show the items and themes we distilled from the raw data. The third part shows which themes are suitable as mechanisms and how these relate to the pivotal points in the timeline of events.

7.2.1 Timeline of events

To explain how DSSU came to be we cannot tell just the story of the idea DSSU itself and its implementation process over the years. Much of the dynamics in this process are in fact a result of processes taking place decades before this. One of the key processes that will be discussed first is the intermittent occurrence of radical changes in the evolution of the Dutch railway system, or at least the sectors intention to instigate these radical changes. Some of the respondents explicitly mentioned such historical processes in order to explain certain tendencies in the sector or problems the sectors were facing during the innovation process. Most notably, respondents explained how one specific improvement program caused the Dutch railway system to divert its development path from the one Japan was seen to have taken. Such a diversion is interesting since the 'Japanization' of the Dutch railway system would mean undoing such historical choice. Hence to grasp the complexities involve in systemically transforming the system towards a new 'Japanese' end-state, as some actors within the sector wished to do, one needs to take into account why the Netherlands and Japan had different end-states to begin with. Therefore we start the timeline of events in the 1960s where the sector decided on this specific program to improve the system.

1960-1985

During the 1960s car transport's increased popularity posed a threat to the viability of rail transport in the Netherlands. The network of highways was expanding while the more capital-intensive railway lines were fixed to some extent. In addition, during the heydays of rail transport in the middle of the 19th century, as we portrayed in Chapter 2, the Netherlands long seemed to rely on water transport as a means to connect larger urban areas. This caused the network, compared to neighboring countries, to be less fine-grained. Whereas cars could bring you directly from point A to B, rail involved many transfers at stations. To cope with the rising popularity of cars, NS started the "Spoor naar 75" (translated as 'Tracks to the year '75') program of which "Spoorslag 70" was the most radical change in the way the system would be operated. Overnight, NS increased the amount of trains that traversed the network by 40%, financed by a new annual subsidy from the government. Even more than a sheer volume increase, the timetable proved to be significant change. From 1970 onwards, NS operated the trains according to a symmetric, clock-face and hub-and-spoke timetable. Also in the routes more point-to-point connections were realized by having trains alternate between destinations, given the same departure point. This new vision on how to operate a railway system had significant impact on the infrastructure. Given the timetables and many different point-to-point connections, central railway stations, such as Utrecht, had to provide for many different connections to be made and if one train would be delayed, other trains that would arrive at the same time would need diversion options. This created railway stations with many railway switches. Although switches constrain the placing of signaling and signaling constrains capacity, the frequency that was used then did not render this a problem. From 1970 onwards, the frequency was two trains per hour for most of the connections.

1985-1995

From 1970 to around 1985 this philosophy helped the railway system to keep up with car transport. However, after 1985 societal developments demanded the need for again a radical overhaul. The expanding economy, environmental concerns and growing suburban cities were triggers for NS to present a new plan for the future of the railways. Based on the expectation that the government would discourage car transport and that a set of infrastructural expansions and more innovative utilization measures would be implemented, the railway system should be able to grow again. This program, Rail 21, in the end proved to be less successful as prior radical programs, such as the aforementioned Spoorslag 70 program. During the implementation of the Rail 21 program, the sector was only able to build the infrastructural expansions and these expansions suffered from high cost increases due to local municipalities demanding sound barriers and tunnels. Also, the attention of the sector slowly shifted from this program to more high-profile investment programs such as the proposed freight-dedicated railway line between Rotterdam and Germany and two high-speed lines connecting Amsterdam with Germany and France. All these developments occurred in an era in which the European Union sought to open up the railway markets and allow foreign entrants to compete with domestic incumbents. As such, in 1995 NS was separated into transport operating company (still NS), and three companies involved in infrastructure maintenance, capacity allocation and traffic control. These three task organizations would later merge into ProRail.

1995-2000

In the years 1995 to 2000 three developments occur that play a role in explaining the DSSU process. Firstly, the ministry (Infrastructure and the Environment, then VROM) appointed Utrecht Central station as the main target of their urban renewal programs. As a central node in the network and an outdated station area, the station would need a significant overhaul to both functionally and esthetically be up to date again. In 1997 the city of Utrecht determined a single design for the station and the surrounding area only to see in subsequent years that communal support was lacking. In 2000 the city and the national government could not come to an agreement and the project was stalled. At the same time, the government was also contemplating on improving suburban rail transport. Many cities, such as Utrecht, had seen increased suburbanization and the growth of commuter towns. The program Randstadspoor was started in 1997 and the ministry placed this program in a larger national investment program so it had a fixed and shielded budget. Randstadspoor aimed to provide more connections between cities surrounding Utrecht and to increase the frequency of trains. In its essence it tries to copy the S-Bahn networks of larger German cities. Accordingly, Utrecht central station will then no longer be a terminus for these regional trains. Something it was for many years.

A third development is the rising interest in Japan as a benchmark. From 1995 onwards, talks about building a high-speed line between Amsterdam and the Belgian border began to solidify and NS tried to both find a way to combine conventional and high-speed rail as well as find ways to turn their railway business into a profitable business. In 1997 a delegation of the product management department of NS and of HST-VEM, the NS subsidiary intended to exploit the high-speed line, visited Japan. The main goal was to study how to effectively combine high speed and conventional rail transport. This visit marks an interesting turning point. For the first time Japan is seen as a viable benchmark that could result in interesting lessons for the Dutch railway system. Besides lessons on high speed rail transport, the representatives see how efficient the Japanese system works and how it provides much more value compared to the Netherlands. Although in the resulting report the lessons do seem somewhat less coherent as later lessons, the scope of lessons is high. The lessons involve preventive maintenance of assets, different ways of designing time-tables, fixed combinations of routes, trains and personnel, dedicated personnel and much redundancy in assets to overcome asset failure. In addition, and this is strikingly different from later visits, is that the scope of differences between the Japanese and the Dutch system is explained by a general theme of designing one's system for one specific function. All changes allow the Japanese system to do one thing: the transport of passengers. And the system is mainly designed to do just that with many sub functions, such as maintenance, being compliant to this main function. Also, the report shows that the lessons are not black and white, that indeed in Japan there are a plethora of differences domestically and that what they do in Japan is simply what works best locally. The report shows, compared to later visits to Japan, a more nuanced picture of why Japan works so well.

2000-2003

In 1999 NS presents the program "Destination Customer" to improve the railway system once again. One of the most crucial aspects of this program is the new personnel planning philosophy where train drivers and conductors will be assigned to fixed routes. Until then, it appeared that complex personnel planning involving mid-shift train changes for train drivers caused many interdependencies between train routes and subsequently a very vulnerable system. Train drivers missing their connection at central nodes of the network caused knock-on delays. Train drivers and unions met the plan with much resistance, as they feared boredom and decreasing autonomy in their work. The controversial plan, although encompassing other measures as well, became known as “Rondje om de kerk” (lit: driving around the church). In 2001 NS implemented this new personnel planning and soon saw how the resistance of personnel was turned into high numbers of absenteeism and eventually large strikes. Punctuality, which the plan intended to increase, plummeted and in 2002 the plan was cancelled.

At the same time, the infrastructure task organization NS Railinfrabeheer was facing problems with maintenance of the railway infrastructure. Railinfrabeheer needed an additional 83 million euros in investments to bring the infrastructure system back to original state. Parallel to these developments internal to the rail sector ran the decision making process of the upgrade of Utrecht Central Station and the surrounding area. After the cancellation of the plans in 2000, new interest arose in the city of Utrecht for revitalizing this area. In 2002, the city held a referendum on this topic asking its citizens whether they preferred a modest upgrade or an ambitious overhaul. The results are that an ambitious overhaul was preferred. As a consequence the city signed a memorandum of intent with the national government about cooperation and co-financing. In addition, urban planners were put to work to design a master plan for the station area and the public transport terminal.

The years from 1999 to 2003 have seen a decline in performance caused by a lack of focus, fragmentation in the sector and unrest in the operational layers of NS. To address these issues urgently, the sector starts the BenB (Benutten en Bouwen, translated as ‘to exploit and to build’) program in 2001, which should deliver a set of immediate actions as well as more long-term actions to deal with the issues the sector is facing. The goal of the program is threefold: firstly, it should be a proof of the sector being able to act as one, and to present an action plan in a coherent manner. This can be seen in the light of increase compartmentalization after the government stipulated split up of NS. Secondly, it should present a set of measures to deal with the decreasing punctuality and the decreasing quality of the assets. Thirdly, it should find ways by which the sector is able to accommodate growing numbers of passengers and freight trains with fewer infrastructural investments. As the program report states the sector should expect an increase in traffic volume of 50% by 2020. Given that the high-speed line and freight dedicated line are being built, and a one billion Euro line between Amsterdam and the Northern provinces is about the be built, the sector realizes that future expansions need to be less costly.

In March 2002, the board of this BenB program visits Japan to see if the country could provide any lessons on how to do this. The representatives stem from NS, the three task organizations and Railion, a freight operating company. Compared to the visit in 1997, the representatives are much higher up the hierarchy and are more diverse. Although the lessons learned are summarized in a report, they do not really take afoot in the organization. Many of the representatives lack the technical know-how to really understand the Japanese makeup of the system and the fact that this visit is taking place in a time where management and operation seem to grow apart, the dispersion of the lessons through lower echelons is minimal. In addition, many of the representatives seem to focus on solutions from Japan that fit their immediate responsibility.

The resulting BenB program encompasses a plethora of measures of which the immediate upkeep of maintenance forms the largest part. The program prioritizes safety and reliability over additional utilization measures and sees building additional infrastructure as the last resort. Robustness is seen as the key word although the word has different connotations than the use of the word in the more systemic approach of Robust Railways years later. Firstly, the sector sees that the entirety of assets needs to be more robust and that the timetable should leave more room for dealing with small asset-related disruption. Secondly, some of the corridor elements of Japan (the separation of different train routes into independent lines) are introduced. From 2007 onwards, the program will instigate the introduction of a simpler timetable with less train routes crossing each other's paths. Although the program does not specifically point to Japan as an inspiration, the resemblances are striking. However, other measures to cope with the increased amount of bottlenecks negate the Japanese principle. Dynamic Traffic Management is seen as way to solve the problems of over utilization of the Schiphol tunnel. In this case, traffic controllers have to locally solve conflicts in the planning by dynamically appointing approaching trains a certain platform.

2003-2007

The period between 2003 and 2007 marks an era with relative stability. Punctuality is again increasing and few real controversies arise around the functioning and governance of the railway system in the Netherlands. Where in 2003 the three task organizations start to become active under the name ProRail, the organizations are officially merged in 2005. In this period there is much contemplation on what each of the sector's party role should be, especially what ProRail as the infrastructure manager should do. This period is also marks the beginning of some extended visits to Japan by someone who would later be seen as the driver of the Japanization of the Dutch railways. In addition this person, together with a fellow coworker, both located in the traffic control department of ProRail establish the Performance Analysis Bureau. This staff department is founded to study actual performances of the Dutch railway network in-depth. In essence, this is in itself an instantiation of Japanese principles, closing the loop between execution (traffic control) and planning (timetable design). Until then, the sector had been organized very sequentially. The way infrastructure is designed and in the end operated follows a linear path from design of tracks and routes, via timetables, to actual operators dispatching and driving trains. The bureau sought to find ways to draw lessons from what actually occurred in the operational layers of the organization and was a reaction to an increased gap between what planners would design and operators would do. This gap arose because planners were used to design timetables freely, only constrained by generally accepted norms and assumptions and operators protected their autonomy and hence decreased the likelihood of valuable evaluations of performance on a daily basis.

In this period a continuing study of Japan by a select group of people with operational knowledge as well as a more in-depth knowledge of what was actually the performance of the Dutch system led to the lessons being learned becoming more fine-grained and coherent, albeit smaller in scope. Firstly, it appeared that asset failure, e.g. trains or tracks breaking down, were only partially the cause of delays. More so, it appeared that the way the Dutch sector planned trains caused delays, especially secondary delays. It appeared that the Dutch system was extremely vulnerable to small delays having a continued effect throughout the network. Parallel to that, but highly interlinked, are the lessons drawn from Japan. Similar to the 1997 visit a more coherent picture of Japan as a benchmark arises. Many of Japan's differences make it perform better, but it is especially in their conjunction that these prove their value. Also, for the first time headway time is seen as crucial. An important discovery is that a difference in a few seconds in minimally allowed headway time results in either a stable or an unstable timetable. It appeared that so-called secondary delays, found to be the main problem in the Dutch network through the PAB analyses, were less of an issue in Japan because of this phenomenon. During delays, when trains traverse the tracks at lower speeds, in Japan the capacity of the network increases whereby in the end the delay automatically has a tendency to decrease. On the other hand, in the Netherlands lower speeds decrease the capacity and create a positive feedback loop and inherent instability. More concrete, the lessons from Japan revolved around optimizing signaling, increasing the speed by which stations can be approached, and an even stronger focus on independent corridors than the BenB program already did.

In itself such lessons were hard to implement giving the sheer cost of all these infrastructural changes. However, at the same time two developments provided an opportunity for this idea to be implemented, regardless of the drivers behind the actual implementation. Firstly, ideas about making Utrecht a throughput station became reality in the 2005 project called VleuGel /Randstadspoor. Besides doubling of tracks around Utrecht to separate regional trains from intercity services and to increase the homogeneity of traffic, the program also intended to create train routes passing through Utrecht, rather than terminating here. In addition, the project will solve the bottleneck in the west of Utrecht where at that time the bridge over the Amsterdam-Rhine Canal only allowed for two tracks. In 2018 many of these infrastructural expansions are to be implemented. Secondly, in 2006 and 2007 a government white paper and a market analysis respectively again stated that by 2020 traffic volume would increase significantly. The sector should expect 50% more passengers and 100% more freight and at the end of 2007 the government puts forward the ambition to introduce so-called timetable-less transport on six of the most important corridors. This would see 12 trains, both regional and intercity services, per hour per direction. In addition the plan includes the separation of freight traffic over fixed freight routes through the country. The ministry of infrastructure and the environment also asks the sector how and at what costs these ambitions can be realized.

2007-2010

In 2008 the sector, ProRail, NS and the representative body of freight train operators start the 'Ruimte op de Rails" program, trying to find a specific way of accommodating increased traffic and the required budget for doing so. Here, two solutions arise. A more traditional approach, involving many infrastructural expansions, would cost 9 to 12 billion euros, a more innovative approach, involving many of the Japanese principles, would cost only 3,2 billion euros. Given that the upgrade of the Schiphol – Lelystad corridor, a project already started and separately budgeted, should be included, the cost raised to a total of 4,5 billion euros. This number had been the result of a few sessions where sector parties together looked for so-called utilization measures on the existing network in a set of expert meetings. This quick scan resulted in a list of measures that would need to be taken if the desire would be to accommodate higher traffic volumes with fewer infrastructural expansions. What was soon decided in these meetings was that Utrecht needed to be overhauled if it were to handle increased traffic on the two corridors that passed through this station. In September 2008 the sector presents the results of the project to the government and in the same month the minister adds the 4.5 billion in the budget for 2009, to be spent between 2013 and 2020 after the Dutch parliament agrees with the approach taken.

The speed by which Utrecht is appointed as a to be overhauled station and the actual contracting out of the work is increased by the pending crisis. In 2008 the Elverding commission intends to increase the decision-making processes of large infrastructural investment projects and the resulting law is adopted by the parliament in 2010. In addition, environmental law is simplified in 2011 and in 2012 the transport minister changes some of the delaying factors of the Elverding committee specifically for DSSU. For instance, additional measures taken to reduce sound pollution are now the responsibility of local governments instead of the national government. Also the official approval of the exploration phase of the project is no longer necessary.

2007 also marks the beginning of the planned timetable changes emanating from the BenB program. Evaluation studies during this period show that the uncoupling of corridors around Utrecht Central Station in the time table had led to an increase of punctuality of 2% point and that there is a decrease of 14% of the number of trains that have to wait before entering the station. However, the study also acknowledges that the initially planned improvements have only been partially achieved. Especially it shows that although in the timetable corridors are separated, planners down the chain not always adhere to this division. Local planners, who in the end make the daily time table, regional traffic controllers and train dispatchers sometimes throw the corridor concept overboard when making ad hoc adjustments to the plan. Last-minute train path requests, for instance by maintenance crews or freight train operators, and train diversions in case of disruptions often cause operational personnel to leave the corridor concept to adhere to other more pressing demands. This observation becomes ever more critical when, in December 2009, heavy snow hits the Netherlands and especially Utrecht Central Station comes to a halt. Although in the timetable, corridors are not expected to be interfering with each other and subsequently a robust and resilient network was expected, this division is very much a theoretical division. Therefore in the first quarter of 2010, the train and traffic control departments of NS and ProRail initiated the project New Action Plan Utrecht (NAU). This project focused on maintaining the corridor principle in the operational layers of the organization as well, especially envisioning the appointment of train dispatchers to fixed corridors. In addition, prescribed contingency plans for disruptions should to a larger extent bear in mind the corridor principle in Utrecht.

In the same period the innovation champion driving the 'Japanization' of the Dutch system increases the frequency of visits to Japan, culminating in a year long stay from the end of 2009. In this period, also managers from lower echelons come to Japan and the lessons from Japan become more detailed, coherent and focused. On top of that, the innovation champion gets the chance to spread his vision on Japan through a series of columns in the organizational newsletter. At his return at the end of 2010, his department, (transport planning) had been involved in collecting the customer requirement specifications (CRS) for Utrecht Central station. 6 months before that, the government had decided to implement the PHS program and had allocated 271 million euros to the overhaul of Utrecht. The end of 2010 then marks the turning point in the Japanization of Utrecht. The transport planning department, drawing the CRS, with the innovation champion in the lead, actively sought to introduce Japanese principles. The multiple visits led them to believe that the key difference was that in Japan main functions of a railway system were prioritized over auxiliary functions, a rediscovery of the lessons learned already in 1997 when NS visited Japan. Now, with a much smaller budget for overhauling infrastructure, purposefully created by ProRail itself, this department had a financial argument for prioritizing the set of CRS to be communicated to the project team. Under normal circumstances, when budgetary constraints were lacking, drawing up the CRS would simply mean collecting all requirements from stakeholders and transferring them to the project team. Now with increased budgetary constraints, conflicts between different requirements could not be simply resolved with expensive civil engineering solutions. Rather, the transport planning department prioritized the set of requirements themselves. For them, this action was a way of taking their role in the organizational network of rail stakeholders: to cautiously use taxpayers' money on infrastructural expansions.

The foundation for the to be communicated list of requirements is a vision by the department on how Utrecht should look like. Inspired heavily by Tokyo central station, Utrecht central station should ideally decrease the amount of switches by 90% and increase the speed by which trains can approach the station. At the end of 2010 the department, after heavy and lengthy internal tinkering, communicates this resulting vision to NS. This created the first conflict revolving around 'Japanizing' Utrecht. The vision, in the form of the future layout of tracks and platforms of the station, is so heavily designed for just the main function that it leaves little opportunity for auxiliary processes to be carried out. In addition, the new infrastructure basically enforces a specific set of routes across Utrecht. For instance, the in 2007 abandoned route Arnhem – the Hague, becomes impossible.

2010-2012

At the same however, December 2010 also sees a repeat of the year before: heavy snow again causes Utrecht Central Station to collapse. Train services are fully cancelled and the system seems to show no signs of any resilience. Even when malfunctions are repaired, the complexity of the node does not allow for a quick recovery of services. Subsequently, the minister demands in January of 2011 that the railway system in general should be less complex, and for the first time, says that it should be more robust. In the end NS comes up with an alternative that still embodies many of the Japanese principles but allows NS to carry out their auxiliary processes as well, albeit it still much less than before. For ProRail, the new vision actually is an improvement to their own idea, since NS solves a bottleneck in the design: the way trains can reach a shunting area south of Utrecht.

The period from 2010 to 2012 is a period best subscribed as increasing compartmentalization of processes and an increasing focus on detail within these compartments. The most integral approach towards Japanization can be found in the Robuust Spoor project where many related issues are studied in their conjunction. Next to that Kort Volgen (and later termed Maatregelen Verkorte Opvolging, translated as 'measures on shortening train succession') studies the use of signal optimization on capacity and safety performance indicators. The latter is especially important since the budget claimed for upgrading the OV-SAAL corridor is calculated in January of 2011 during the detailing phase of PHS and the assumption is that signal optimization would increase the capacity of the corridor without investing in fully doubling the tracks. From then on, signal optimization becomes part of PHS.

Robuust Spoor (RS) embodies most of the notions found during the period the sector studied Japan but remained relatively abstract, especially compared to the developments at DSSU. Although DSSU uses the principles from RS through the involvement of the initial innovation champion and the transport planning department, DSSU seems to be running parallel to the RS project. RS intends to upgrade the entire Dutch network to increase the robustness and appoints 50 projects as key projects in the near future. At these projects, medium-scale improvement projects of existing nodes in the network, the use of RS principles would mean less costly projects and increasing the networks ability to cope with small delays.

The project RS can be seen in light of the sector's wish for a more collaborative and integral approach towards innovation. It is broader in scope and scale compared to the initial design for DSSU. In that sense, it began as a more systemic innovation than what DSSU purported to be. However in June 2011 ProRail holds a meeting to discuss the progress of the RS project and the apparent lack of a clear vision. In subsequent meetings that month, the project sees that coherence is lacking in the 50 projects and that it is not clear which transition approach should be used. ProRail therefore decides to start of so-called pre-project Robuust Spoor in order to dry-test the implementation of RS principles at an actual top 50 project. The project is done without any involvement of outside stakeholders.

In September 2011 the results of this pre-project indicate that RS principles are closely linked to Kort Volgen principles. In redesigning the corridor Hoofddorp – Amsterdam, with Schiphol Airport in between, the distance between the signals was too much of a constraint. Hence, one often will see that implementing Japanese principles will entail both corridor separation and signal optimization and that the implementation of the one forces the implementation of the other. Additionally the project members see the Japanese principles as guiding but find the lack of operationalization and the lack of clear performance indicators ineffective. Also second-order effects of the implementation on other processes, such as maintenance and the planning of rolling stock and personnel are topics for further research.

Parallel to the study project on implementing Japanese principles, DSSU and two other projects are instigated. Firstly, for the DSSU case, the transport planning department specifies the customer requirement specifications and communicates these to the project team. This occurs concurrently to the RS and KV projects. They find an engineering consultancy to make a plan adhering to these CRS. Secondly, project NAU in more detail looks at applying the corridor concept in the traffic control department. Thirdly, decreasing headway times by optimizing signaling becomes part of a larger program looking at multiple ways by which the succession of trains can be increased. Although the three streams tackle, partially, concepts that have originated from Japan, their coordination is only limited. In addition, whereas the latter project remains more of a research project, the other two also involve direct implementation. Especially, the DSSU project has a high amount of irreversibility, which will explain the dynamics found later in this period (until 2013). Figure 7.3 graphically depicts the divergence of different projects from the same origin.

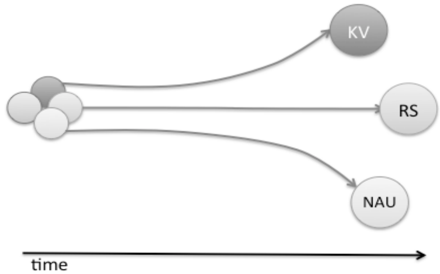


Figure 7.3: from a coherent set of changes, the initial idea disperses over different projects (KV, Kort Volgen: signal optimization; RS, Robuust spoor: corridor separation in infrastructure; NAU, Nieuw Actieplan Utrecht: corridor separation in traffic control) and loses its systemicity

The NAU project, started in 2010, comes to fruition in the first quarter of 2011. In February of that year, the project team conducts a gaming simulation experiment where the separation of corridors and the new division of work is tested. Actual operators of both NS and ProRail play a range of scenarios involving realistic disruptions on a realistic depiction of the Utrecht infrastructure and with trains moving according to a realistic timetable. In this game, many switches are removed from the board and the project team is interested in the extent to which traffic controllers are still able to cope with disruptions. The results of the game are controversial. For some disruptions, the corridor concept will lead to complete train routes being cancelled, whereas in the then current situation, diversion possibilities would be present. Also, the corridor concept demands from traffic control a much more proactive approach. Proactivity meaning that trains have to be controlled far beyond the boundaries of Utrecht to manage Utrecht itself. This is controversial because it would mean a shift in responsibilities, expected to go from the Utrecht echelon to the national traffic control echelon. The fact that this echelon was only recently installed (in 2010) added to the controversy.

A month later the project team starts the actual pilot of NAU, limiting traffic controllers of Utrecht in the range of tracks they can use of controlling traffic. The results of the game had led to a big discussion on how to deal with disruptions and who should be responsible but these discussions had not led to timely solutions before the start of the pilot. However, the game also showed that some specific disruption scenarios were unsolvable if the corridor concept would be strictly adhered to. Therefore, the clients of the project allow for three exceptions under which traffic controllers are to leave the corridor concept.

In July 2011 the results of the pilot are promising enough to make the steering committee decide to implement NAU as a basic principle in redesigning the traffic control of large hubs of the Dutch network. However, the traffic control departments have also seen the negative side effects of a separation into corridors. Until now, with the railway switches still physically present, they could decide on exceptions to the rule.

At the same time, the DSSU design more and more converges on a single solution. After the conflicts between NS and ProRail at the beginning of 2011, in March a special meeting between the ministry, the city of Utrecht, Ns and ProRail results in the decision to build a specific design of DSSU. The design is slightly suboptimal regarding robustness, but is less expensive. The main feature of this design is that on the south side of Utrecht the 8 lines are divided into brackets of 4 and that the shunting trains on platform 8 and 9 now still have to cross a corridor to reach the shunting area.

Whereas the overall design converges to one specific layout of tracks, the project team is involved in making more detailed designs on the new station. With the very strict CRS in mind, and the new prioritization of main functions over sub functions, they have difficulties to design something that adheres to these requirements. Especially the desired headway time of two minutes seems to be problematic. Given that a buffer is always demanded, the technical headway time should be 90 seconds around the station. Theoretically, this seems already impossible, but to still approach this value they start to implement signal optimization on large parts of the tracks leading to Utrecht. According to the project lead, this in itself is not forbidden by the design protocols of ProRail and the method is already used in 83 other locations in the Netherlands. However, the scale by which it is used and especially the repeated use over the same route is new and highly controversial. At the same time, this measure is studied as one of the ways by which the succession of trains can be sped up in an entirely different project: Kort Volgen. Also other measures are taken that according to the prescribed design protocols are not desirable. All these deviations of the protocols, the project team has to communicate to the train safety department of ProRail. During 2011 and 2012 this department sees an extreme influx of these deviations. Whereas normally they see 40 deviations per year for all projects in the Netherlands, DSSU alone results in the same number. This department in the end halts the process because of two reasons. Firstly, the repeated use of signal optimization worries them as this can have serious safety-related consequences for the driving behavior of train drivers. Secondly, the emergent effects of 40 interrelated deviations of the guidelines are hard to understand for the department. In their conjunction, multiple deviations can lead to unforeseeable safety effects.

Parallel to this design process runs the process of awarding the franchise for the main part of the Netherlands to NS by 2015. Given the ministries desire to start the PHS program as early as possible, NS is putting pressure on ProRail to accommodate frequency increases on the most important corridor of the network. NS actively seeks ways to start the PHS program on the A2 corridor and asks ProRail if by 2015 Utrecht would be able to accommodate this. A study by ProRail finished in October 2012 however shows that this is impossible. Most of the features of Utrecht needed for this are only finished well after this date.

2012-2014

In 2012 also two other stakeholders in the process start to get worried. The logistics department of NS and Nedtrain see the side effects DSSU has on their processes and they fear an increase in operating costs. For Nedtrain it involved the reachability of their maintenance area on the north side of Utrecht caused by a higher frequency of trains and limited amount of switches. The logistics department sees that operating cost reducing measures, such as the changing the size of trains during the day, become impossible. Also, and for the overall process an argument that is more widely shared, they point to the projects design being based on an hourly timetable. DSSU is optimized to accommodate a specific timetable, but NS might not use this timetable all the time. In the off-peak hours and during weekend a more thinned-down timetable might be more economically efficient. However, the way DSSU is designed might actually create a decrease in performance. Additionally, the department sees problems when, given this design, the organization has to start services in the morning and end services in the evening. The two departments together start a simulation study to study the actual effects of the design of DSSU on their processes. In July 2012 results show that NedTrain can still service trains, but that the timing becomes very critical. For the logistics department the results show that changing train size during operations becomes nearly impossible.

Of all the worries, the worries about the signal optimization and its effect on train drivers are the most obvious at that time. NS formally requests ProRail to study this and in august of 2012 the NLR is tasked with studying the effects.

In 2012 the ministry informs the parliament about the progress on PHS. In the letter they send, there are the first signs of worries about the DSSU project. The Randstadspoor project, started in 2005, intends to double the tracks between Utrecht and the commuter town of Houten and DSSU highly depends on this. However, they foresee problems with the permits needed to commission this new part of the infrastructure. Additionally, the ministry points to the risks of using Japanese principles in the design of DSSU as the concrete specifications of these solutions and their timely delivery are not guaranteed.

Also in 2011 and 2012 building work starts at the OV-terminal and the adjacent shopping mall. With the DSSU works scheduled to start in 2013, the complexity of the overhaul increases. On a relatively limited area, three large projects will be carried out, by different contractors and with different clients. All this work has to be done in small pieces as not to interfere too much with the ongoing flows of trains and travellers. However NS, who has to accept the specific timing and periods of work, officially refuse the proposal by the capacity planning department of ProRail.

Three conflicts result in a series of workshops at the end of 2012. The use of signal optimization, the timing of the start of PHS and the planning for the construction works are topics to be discussed. Of the three the signal optimization issue seems to be the most pregnant and the initiators of the workshops tackle this one first. The core of the problem here is that the design team had differently adopted the design protocols than other stakeholders had expected. Although the use of signal optimization is not strictly prohibited the extensive use over longer ranges of the track is unprecedented. In the eyes of the lead engineer of the project this did not warrant additional research, but other departments focusing on safety issues did not agree. Given that they had to commission the design and use of the infrastructure they were in power to demand changes. In the end the workshop decided on conducting an extensive simulator test with the new signaling regime and await the results before granting the contractors permission to build the new infrastructure.

Another issue is that within the Dutch railways there is a desire to start the PHS timetable on the A2 corridor in 2015. In 2015 the Dutch government would extend the franchise agreement for another 10 years and a quick win for the Dutch railways to be realized early on is the increase of frequency in trains over the Netherlands' most densely used corridor. On the other hand, departments within Dutch Railways tasked with maintaining the current timetable during construction works see ProRail intends to temporarily decommission the infrastructure in 2014 and 2015 for unacceptable long periods of time. Here Prorail is faced with conflicting demands from their main counterpart. Having the infrastructure be ready for increased frequencies earlier on will, if it is at all technically feasible, cause construction works to be more densely located in time. This will stipulate the use of entire weeks and many weekends to do the needed work and will be a severe hindrance to the final customers. For the logistics department of the Dutch Railways this is unacceptable. In the end this conflict is resolved by forgoing on the early introduction of PHS and the use of one 9 days period and a weekend for the main part of the renovation works.

In the spring of 2013 the innovation department of ProRail conducts a simulator test where train drivers have to drive over tracks with and without signal optimization. The results indicate that critical safety issues are not to be expected but that these findings only hold for Utrecht and no nation-wide lessons can be drawn.

2014-2016
Immediately after the results indicate that it is safe to use the new signaling regime around Utrecht central station the minister of infrastructure gives a go to the building of the project and the contracts are signed with the contractors. From here the design of the infrastructure only barely changes. Most controversial issues from this time on involve legal issues such as problems with the needed permits. In 2014 there are problems with the permits for the related Randstadspoor project and causes a year delay for the DSSU construction works, as well as a budget overrun of 30 million euros. As of now, construction work is still underway and the scope of the project has been set by the ministry to avoid further budget overruns and delays. Regarding the design, the stakeholders involved or the institutional setting there

are no changes occurring anymore as well as not to be expected before final commissioning in 2016.

7.2.2 Pivotal points

The timeline shows a myriad of processes that have taken place in parallel and influenced each other simultaneously. For instance, both the development of the idea of 'Japanization' and the idea of overhauling Utrecht central station took place at the end from the 1990s to around 2010 while only at later stages these processes started to interact. The connection of the innovation process and the station redevelopment process therefore was a pivotal point since it allowed the innovation to find a window-of-opportunity for its implementation. To better discipline the analysis of underlying mechanisms we summarize the timeline using more of these so-called pivotal points:

1. The notion of Japan as a serious suitable benchmark started in 1997 and the idea of Japanization became more and more operationalized.
2. Early implementations of elements of Japan seemed to be suboptimal or even controversial; this did not impact the viability of Japan as an idea.
3. The founding of the Performance Analysis Bureau seems to be crucial in understanding the operationalization of the Japan idea from 2005 to 2010.
4. The push for using less capital for adhering to the PHS demands from the national government by elements within the railway sector seemed to implicate the use of Japanese principles.
5. Several projects were windows-of-opportunity but only Utrecht seemed to be effective after the first push for Japanization.
6. The ramifications of using Japanese principles in redesigning Utrecht became known only at later stages to those stakeholders not fully involved in the design process.

If we look at how the Japan idea developed from 1997 to its implementation at DSSU given the changes in the context, we arrive at Figure 7.4:

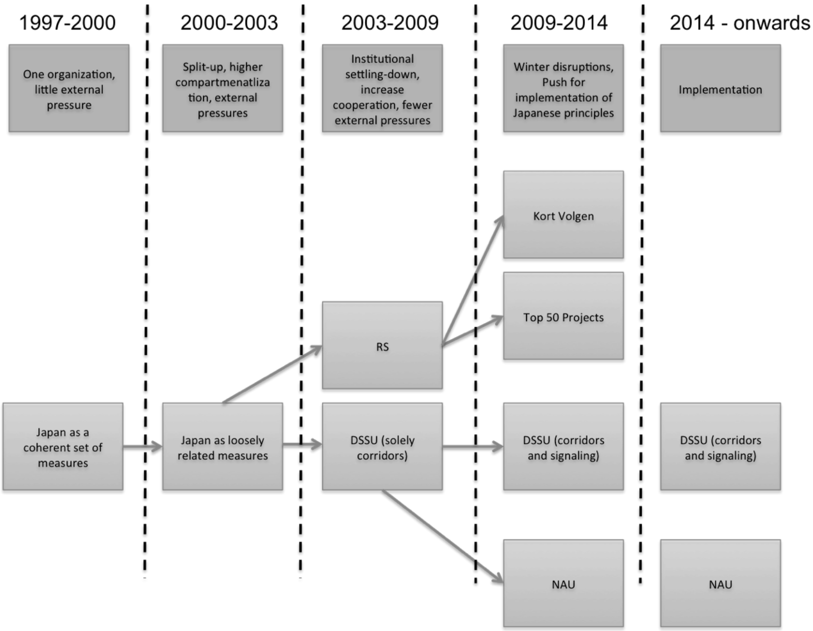


Figure 7.4. Pivotal points over the course of the innovation process

7.2.3 Item selection from raw data

Next to insights on what had happened during the process of implementing the Japanese principles, respondents also provided insight on why the events described previously happened. This allowed us to more abstractly explain certain recurring mechanisms. To do this, we partially transcribed the 25 interviews. From this we distilled items, which we used to form common themes.

From the transcribed interviews we started to collect interesting items per interview and built a list of items based on the first 15 interviews. We reached theoretical saturation after the 15th interview, meaning that no additional items could be found that could significantly contribute to theory building. The remaining 10 interviews however did help in making the theoretical understanding of the case more specific as well as provide us with additional examples for the mechanisms we ultimately found. We first provide an overview of the items we found in Table 7.1.

Table 7.1. Grouping of items recovered from interview and document data.

Topic	Items	Notes
System	Complexity, rigidity Many stakeholders Different incentives Interdependence	Respondents provided general remarks on the railway system as to being complex and rigid and how this complicates innovation processes.
Institutional setting	Compartmentalization Organized per aspect system Coordination through rules and norms between aspect system Coordination through communication and redundancy within aspect system	Mostly using historical explanations, many respondents found that compartmentalization per aspect system increased the effectiveness and autonomy for departments but that this sometimes came at the cost of local optimization. Coordination between compartments was mainly pre-arranged through the use of rules, norms and standards. This however causes little ad-hoc communication between departments and between operational, tactical and strategic echelons.
Innovation	From abstract theory to workable artifact, devil is in the detail Innovation crosses original organizational and departmental boundaries Innovation is a local optimization Innovation has far-reaching effects on other processes	Respondents especially noticed the fact that on abstract levels the innovation is not that contested but that when design processes reach the micro-level detail phase, dilemmas appear. Additionally they described the innovation as systemic and multi-dimensional, crossing many aspect systems, but also solely focused on one local area (Utrecht). This in juxtaposition to the standard of network-wide optimization per aspect system.
Project dynamics	Increased interdependence between projects Project interferes with ongoing operations External influence on project Design of innovation constantly changes	Besides a mere account of the technical reality of the innovation, the interview data also encompassed notions on how the project around the innovation was increasingly influenced by dynamics in other projects. This led to the shared observation that the innovation itself was constantly changing and made it hard for stakeholders to come to workable agreements. Two respondents mentioned the sectors reliance on large-scale programs to diminish the influence of external pressure and overcome the problems of lack of attention and the rigidity of the system but added that this usually involves the collecting of several smaller-scale projects into one.
Assumptions	Use of assumptions to guarantee progress Assumptions have their own dynamics	Especially respondents not directly involved in the design process of the innovation mentioned how assumptions were used to ensure progress but that these assumptions also were never really documented or taken into account later on.
Transparency	Innovation process is a black-box Effects of innovation are unknown	External stakeholders found it hard to know exactly what was going on in the innovation process, what the innovation truly meant for them and in what direction the project was pushed by the initial innovation champions.
Shared frames of references	Different languages Different interpretation of research findings Different definition of innovation	A consequence of compartmentalization is that when different departments have to cooperate, they stem from different historical background and have developed their own language and own frame of reference.
Dilemmas	Dilemmas are not observed Dilemmas are not tackled No clear process for escalation Dilemmas halt progress	All respondents from the initial 15 interviews mentioned that the core of the problem is that dilemmas are encountered during the innovation process but that the sector finds it hard to accurately and effectively deal with these dilemmas. Given that dilemmas are inherently unsolvable, the technical-oriented sector has little process agreements on how to cope with these dilemmas. The dilemmas usually come to surface with a time lag, caused by those in favor of progress, such as the innovation project team not communicating potential dilemmas to external stakeholders.

7.3 Analysis

From the set of items we tried to distill a more abstract theory that could explain the pattern we found in Chapter 6. We noted that the respondents explicitly mentioned the trade-off between project progress and the tackling of dilemmas upfront and that this strongly overlaps the notion of suppressed volatility to the detriment of the so-called fuzzy back-end of systemic innovation processes. We see that those looking for progress, such as the innovation champion, the transport planning department and the project team have a trade-off between signaling dilemmas and ensuring that the project moves on. When dilemmas do come to surface the institutional setting and the behavioral tendencies of stakeholders usually cause a delay. Because this cluster of items of dilemmas and progress fit perfectly the notion of volatility at later stages, we used this cluster as an anchor for our further analysis. In the set of remaining items we looked for ways to explain why dilemmas pop up in later stages, where these dilemmas come from and in what ways those involved in the process could postpone the tackling of these dilemmas.

From the list of items we distilled clusters of themes. These themes comprise of a set of items that were covered by multiple respondents. Whereas the items stay close to what the respondents actually were saying, the grouping of these items into themes was partially disciplined by the items themselves and partially determined by the freedom of the researcher.

The analysis of the items and the grouping of them into themes resulted in nine themes that either depicted static context phenomena (4) or depicted dynamic mechanisms (4) resulting from the static context and the introduction of a systemic innovation (1). By static we mean phenomena that are facts of life with which stakeholders directly and indirectly involved in the process have to deal with. They are also static in the sense that they are only to the slightest extent controllable by single actors, even though some of them involve their own behavioral tendencies. The dynamic mechanisms are a result of these static contexts combined with the trigger for a systemic innovation. These mechanisms basically show what happens if we let loose a systemic innovation on a static model that is represented by the four static context phenomena and see the dynamic behavior of this model over time. The analysis results in two general macro-level patterns that describe the positive and negative effects of the four micro-level dynamic mechanisms found. These set of themes were then used to analyze all 25 interviews. Table 7.2 provides a summary of 25 interviews using the themes

Table 7.2 Grouping of items into themes and their prevalence in interview data

Theme	Name	Description	Prevalence (out of 25 respondent, the amount of respondents mentioning this theme)
A1	Systemic innovation	Comments on what makes innovation special and in what ways this is noticed	7
B1	Dynamic environment	Comments on the environment (political pressure, consumer pressure)	17
B2	Path dependent system	Comments on the railway system as complex and inert	12
B3	Professional fragmentation	Comments on compartmentalization within the sector	16
B4	Substantive coping strategies	Comments on how actors deal with complexity	18
C1	Ambiguity	Comments on the ambiguous nature of the innovation and the inability to predict the consequences of implementing such an innovation	20
C2	Disentanglement	Comments on how innovation was operationalized from 'Japan' to something implementable in the Netherlands	11
C3	Interlocking	Comment on how the innovation interlocked with other projects and how subsequently these projects combined with other projects	17
C4	Shielding	Comments on the level of inclusion of the innovation stakeholder arena and the permeability of project boundaries regarding information and communication	17
D1	Progress	Comments on how the project moved forward	13
D2	Dilemmas	Comments on the emergence of dilemmas and the sector's problems with coping with these dilemmas	16

With these themes we have covered most of the respondents and most of the themes have been touched upon in equal amount. One outlier is theme C2 about the gradual operationalization of Japanese principles to make it workable for the Dutch system. This has to do with the fact that only those respondents that were involved in the process of instigating the innovation (the innovation champion and the project team) were able to provide insight into this mechanism. Outside stakeholders that were only involved later on were less able to comment on how this mechanism had taken shape. In Figure 7.5 we present an overview of the 11 themes we have found and show their relation. In the next paragraphs we will discuss each theme its relation to other themes according to the analysis.

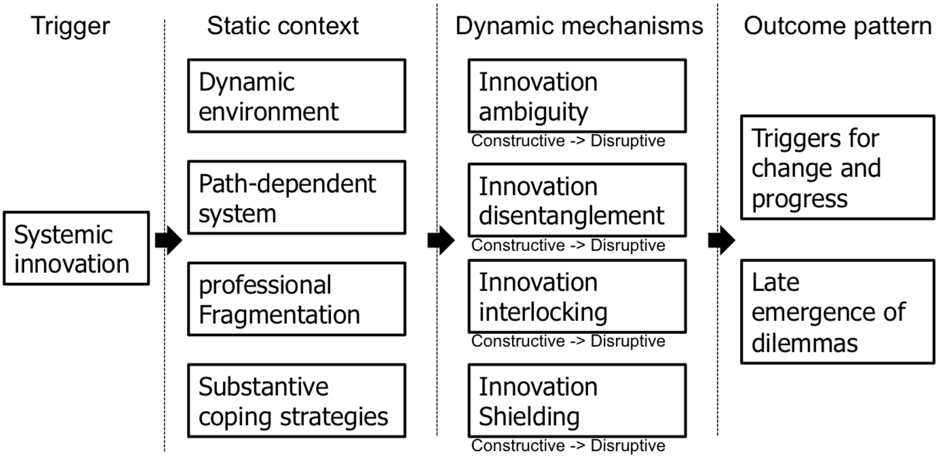


Figure 7.5. Static context and dynamic mechanisms

7.3.1 Systemic innovation

The initial focus of this chapter was the implementation of so-called robust principles at the rebuilding of Utrecht Central station but the analysis showed that these principles were just a subset of a larger set of changes. What these changes had in common was that they stemmed from Japan as a benchmark. Our analysis starts in 1997 as the Dutch Railways first visits the country to study how the country is able to outperform the Dutch system on indicators such as punctuality and economic efficiency. The document detailing the lessons learned already hints to the systemic qualities of the difference between Japan and the Netherlands, and shying away from explaining the difference due to technical differences. The name of the document ‘simply perfect through perfect simplicity’ already shows that Japan is systemically different and that all these differences can be explained by the continued focus on simplicity rather than flexibility, a focus the Dutch network had sought since the Spoorslag 75 program. The detailed lessons in addition show that Japan’s performance has no single explanatory variable but is more an emergent outcome of many contributing factors, with the general theme being its focus on optimizing one single transport function. Summarizing Japan seemed to be different on many aspects:

1. Highly reliable assets, redundancy in assets
2. Disciplined operational staff
3. Fixed combinations of infrastructure, rolling stock and personnel
4. Integrality in the strategic and tactical layers of the organization
5. Stations optimized for easy transfer
6. Lean traffic control and dispatching regimes.

Later visits corroborated the finding that no single explanation can be found for Japans’ high performance but add to that the notion of its systemic nature. Additional differences found during these visits encompassed the separation of the network into independent corridors, optimizing the placement of signaling, the removal of railway switches and a high fit between the infrastructure and the timetable of trains. These changes can be called systemic for three reasons: firstly, they are all dependent on each other in realizing to overall edge the Japanese railway system has, and simply implementing a few of them in the Netherlands will not render the expected results. Secondly, these changes involve a range of qualitatively different parts of the system, such as technical artifacts, design principles, procedures, cultures, and organizational setups. Thirdly, as apparent later on during the actual process of implementing it, introducing Japanese elements in the Netherlands created many cascade effects.

7.3.2 Dynamic environment

When talking about an environment we have to first delineate the system around which the environment is located. In the previous chapter this delineation took the form of the technical artifact and the social actors and their institutions that directly impact the technical makeup of the system, i.e. the project. Hence, we see those actors that demand something from the system yet cannot directly impact the makeup of the system as the environment. These demands result in pressures on those directly involved in designing and operating the system, as they are responsible for accommodating the demands. This is crucial since systems survive as long as they receive energy from the environment (materials, finance, political support). To ensure this, the system has to perform functions that the environment appreciates. Concretely, relevant parts of the environment in this case comprise of passengers, public organizations and transport authorities, ministries and transport-related pressure groups.
The analysis showed two defining characteristics of the environment that were relevant in explaining the resulting dynamic patterns. These are its inherent multi-dimensionality and its erratic nature.

Pressures from the environment on the system can take on many different forms and stem from a range of diverse stakeholders. Their nature can be quite different, as passengers are a more heterogenous group with a range of demands communicated mainly through their willingness to pay, whereas governmental bodies usually have range of direct communication channels available. The nature of these combined demands comprise of many different dimensions such as economic efficiency, safety, capacity, speed, punctuality, noise pollution, and the amount of direct connections available. Two problems occur when a system is confronted with a range of different demands: firstly, some of the demands are inherently conflicting. For instance, passengers that traverse a certain route desire little to no transfers, demanding train services to halt at every station; also desire short travel times, demanding a limited amount of stops at stations. Other usual conflicts between demands are high capacity utilization and economic efficiency and safety, as safety measures cost money and limit the use of capacity. Also, higher capacity utilizations make the network more economic efficient and provide passengers with increased frequencies but also make the network prone to delays spreading through the network, decreasing punctuality and robustness. Secondly, some of these pressures are easy to quantify and measure, whereas others are more subjective, only available in qualitative terms and hard to measure. Capacity is, to a certain extent, measurable in simple numerical figures whereas safety or travel comfort is harder to quantify. In dealing with dilemmas that juxtapose pressures with different measurability, this proofs to be problematic.

Next to how these pressures look like and how they can or cannot be measured, many respondents deem the erratic nature of these pressures important. Erratic here means that the change in pressure cannot be predicted based on previous trends, and hence are hard to cope with. This is a distinction from more predictable changes in pressure. For instance the sector is confronted with an increased focus on cutting costs and becoming more economically efficient. However, even such more predictable trends suffer from sudden changes such as when elections bring about new perspectives on public transport. Especially swifts between right and left wing coalitions mean respectively decreased and increased funding for public transport. More erratic are sudden demands for higher safety after accidents and sudden demands for robustness after large-scale disruptions. It is noticeable that values which can be put into political pressure easily are more erratic then pressures that stem from a large and unrepresented crowd of actors.

7.3.3 Path-dependent system

Many of the problems actors within the railway sector have to deal with are related to the fact that they need to accommodate multi-dimensional and erratic environmental demands with a system that is highly inert and path-dependent. Firstly, high capital intensity and durability of assets mean that bringing about changes in the technical makeup of the system need to take into account long payback periods as well as only few windows-of-opportunity when assets are being replaced. Secondly, the interdependencies between constituent elements of the system create complex relations between the makeup and its functional performance. This creates many conflicting functions such as a meshed network and a speedy network, which one cannot have at the same time. In addition, the constant adaptation between system elements over time creates the phenomena of path dependence and technological trajectories. The current problems the sector is facing is, according to some, the result of historical decisions to keep freight traffic and passenger traffic combined and to allow for a very specific time table setup in the 1970s. The consequence was that station areas became hubs and needed many railway switches to allow for the arrival of many trains at the same time. Hence, from there on, the Dutch railway system evolved quite differently from the Japanese system which favored simplicity and optimization of local lines over flexibility and optimization of national networks.

7.3.4 Professional fragmentation

Whereas the technical system is made up of highly interdependent elements, the social and institutional setup around it is one of high separation and little interdependence. Many respondents have noted a high degree of compartmentalization in the sector, most notably arranged according to profession and discipline. A careful look at the ways this fragmentation manifests itself shows that fragmentation is present along two dimensions: horizontally and vertically. Horizontal fragmentation refers to a clear boundary, mostly in communication, between strategic, tactical and operational decision makers. Vertical fragmentation refers to the boundary between different disciplines involved in the railway system, such as capacity planning, traffic control, safety engineering, asset management, and civil engineering.

Horizontal fragmentation is seen as a remnant of the history of railway systems as it used, and still is, organized very militaristically. Plans were designed in higher echelons of the sector and, lacking proper communication tools, operators on the ground were expected to follow the plans exactly. This way of organizing still persists, although initiatives have been undertaken to diminish the fragmentation to some extent. Many have contributed this feature of the sector to the desire for autonomy. Especially in lower echelons of the organization, respondents have pointed to the lack of willingness for thorough evaluation of daily performance. This is especially problematic if one intends to bring information back from operational layers of the organization to more tactical and strategic layers.

Vertical fragmentation has a somewhat similar origin as horizontal fragmentation, but mainly can be explained through the system's functioning as a machine bureaucracy. The operating of the current system, and the design of additional infrastructure closely follows a linear model. Usually, infrastructure is designed first with the overhead wiring as the lead function. Next to that, switches are placed and safety signaling is designed. When the infrastructure itself is designed, other departments design suitable line routes and subsequently another department designs the exact timetable.

In some instance this fragmentation overrules organizational boundaries, as some respondents have shown how cooperation with outside parties from the same discipline is much easier than cooperation with internal departments from other disciplines. For instance, in designing timetables there is a close cooperation between the capacity planning department of ProRail and the logistical department of NS, and advances have been made to increase the cooperation even further. However, when cooperation is instigated with departments from other disciplines, even within ProRail, problems arise to a larger extent.

Fragmentation is a fact-of-life with which actors in the sector have to deal. However, fragmentation has its advantages and there are many factors explaining why to some extent this proves to be effective. Firstly, it gives each organizational entity a relative amount of autonomy. Hereby they can respond to environmental pressures relevant to them in a certain unrestricted way, without having to take into account the boundaries with other disciplines. Secondly, some functions enjoy scale and network economies. Infrastructure that consists of the same type of infrastructural element throughout the country is for instance much cheaper and easier to maintain and to repair. Hence, the asset management department states what set of elements can be used, disregarding potential local optimization. The same holds for safety systems, where higher safety is ensured when safety regimes are homogenous throughout the country. Timetabling enjoys the same network economies, where network coherence means optimal travel times, transfers and connections. Network-wide optimization per aspect system (safety, timetabling, etc.) is, through this fragmentation, possible.

7.3.5 Substantive coping strategies

A fourth static context theme that appeared from the data focuses on the way actors in the railway sector deal with complexity and uncertainty. Decision makers, designers and project managers are dealing with complex issues related to a complex technical system that needs to provide for multiple functions.

Complexity in this case involves a plethora of dimensions. The data shows that unknown parameters or unknown parameter values is one of the main reasons that designing is complex. In addition, complexity involves time pressure and dilemmatic choices where functions appear to contradict each other. In dealing with them, the data shows that there is a tendency for more substantive strategies to be applied, that is strategies are used that focus on solving the complex problem substantially, rather than using more process-like or organizational ways of dealing with complexity.

Three ways of dealing with complexity appear to be used in most instances, and these close relate to the aforementioned fragmentation in the sector. Firstly, there is a tendency to reduce complexity substantially by delving deeper into the problem and finding ways out of this complexity by additional research. This allows actors to uncover parameters or fix certain parameter values. Secondly, another way to ensure progress in the project as well as deal with complexity and uncertainty is by using assumptions. Whereas the first uses time to solve complexity, the second one uses assumptions to temporarily solve complexity. Here there is a clear juxtaposition between complexity reduction and progress. A third one is a strategy of simply neglecting complexity and to some extent resembles the strategy of assumptions in allowing progress, but the forgo on the explicit use of assumptions. That is, with both strategies complexity is temporarily reduced but with the strategy of neglecting there is little awareness of this temporality.

The data shows another coping strategy that is deemed relevant in dealing with dilemmatic complexity and that is using escalation mechanisms where additional actors are introduced into the decision-making arena. Usually these actors stem from higher echelons in the organization and they have to decide on fixing parameters or trading-off the pros and cons of certain dilemmatic choices. Respondents have noticed that there is reluctance within the sector to adopt this specific strategy. This is because escalating to higher echelons is expected to increase the uncertainty about the final solution and decreases the power of lower level echelons.

The focus on more substantial coping strategies can be explained through the high professional fragmentation. When fragments of an organization are highly separated, involving new fragments to a decision making process will decrease the predictability of the process for the incumbent actors.

7.3.6 Contingency

The aforementioned static qualities of the environment in which a systemic innovation process takes places is highly contingent on the actual challenges that had faced the railway sector. That is, although they may proof to be disruptive or counterproductive in some instance, overall the setup of these qualities was workable. Given a highly complex and inert system and a dynamic and multi-faceted environment, some fragmentation allowed stakeholders to cope with pressing issues in a quick manner and without too much organizational coordination. All this apparently worked well, and as some respondents noted, worked well because especially capacity utilization was low and financial resources were abundant. Faulty assumptions about operator behavior or local conditions in network when designing timetables proved to be only causing minor problems later on. Departments working on detailing the timetable or actually carrying them out still had many degrees of freedom to work around the errors in the initial plan. In other words, under conditions of abundant financial resources and lower utilization of existing capacity, the complexity of the system is lower and hence warrants the ways the sector is organized and the ways actors deal with complexity as effective.

7.3.7 Dynamic mechanisms

However, the analysis shows that the static qualities also backfire when a sector is dealing with implementing systemic change. In this paragraph we describe four general mechanisms that emerge from the data. These mechanisms arise when a system, statically described in the aforementioned four qualities, is confronted with the trigger for a systemic innovation. As will be shown, some mechanisms make it easier to effectuate such an innovation but they later on play a large role in problems uncovered over time. The general mechanisms that emerged out of the data deal with the ambiguity and 'systemness' of the innovation artifact itself, and the interlocking and shielding of the actor arena around it. In this part of the chapter we will describe all four mechanisms and show their interaction effects during an innovation process. Since these mechanisms inherently deal with dynamic phenomena we provide graphical depictions of these mechanisms that show what these mechanisms mean when an innovation as an idea moves towards implementation into a concrete artifact by a collection of organizational entities. We found that each dynamic mechanism can be explained using the depiction shown in Figure 7.6

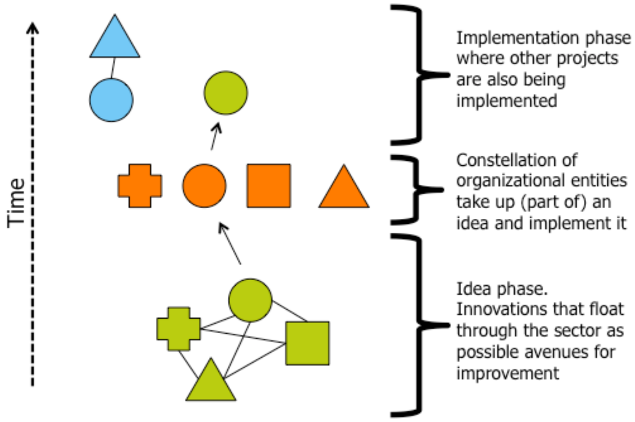


Figure 7.6 General depiction of a systemic innovation process

In this depiction we see a systemic innovation (green) that is partly taken up by a constellation of organizational entities (orange) and sees parallel implementation processes of other innovations (blue) with which it can interact. In this specific example one department (orange circle) implements a part of the innovation for which it is responsible (green circle) while neglecting other parts of the innovation (e.g. green square) and the inherent interdependency between the innovation elements (notice the links between the green objects)

7.3.7.1 Mechanism 1: innovation ambiguity

In general, multiple respondents pointed to the fact that no general consensus was present on what actually constituted the innovation and what the functionality of it was. From the data emerges a picture of ambiguity related to the innovation artifact itself. This ambiguity first and foremost is related to the innovation not being one concrete artifact itself but more a collection of related smaller scale changes. Hence, what was considered part of the innovation changed over time as well as what it intended to change within the incumbent technical system. We see that this phenomenon shifted from being constructive in the beginning towards being disruptive at later stages of the process.

Constructive ambiguity relates to the malleability of the innovation in both the technical makeup of the innovation and what it should bring about when implemented. In earlier stages, the innovation was simply considered everything they did differently in Japan and made the Japanese system outperform the Dutch system. In 1997 the study visit delivered a set of lessons about why Japan was able to accommodate higher traffic volumes, against lower costs and with higher punctuality figures. The coherence of all these related differences was acknowledged, although in later years this coherence decreased. It caused the innovation to be a panacea for everything, since each element in itself, such as preventive maintenance, could be noteworthy for the Dutch sector. This especially holds given the static context described earlier. When fragmented organizational entities have to cope with a rigid and complex system and dynamic multi-faceted pressures from the environment, the specific configuration of the innovation at that time allowed the sector to cherry-pick elements of Japan. While the coherent implementation of it was endangered, and at that time yet to be discussed, it allowed the idea of Japan in general to persist for prolonged periods of time. Even when only very little actually was implemented, the idea Japan as a direction to go, at least on the level of fragmented departments, remained alive. Even major setbacks such as the controversy surrounding the implementation of fixing personnel to train routes, did not cause an entire cancellation of the idea of Japan. Potentially, although the data does not show this, the removal of implementable parts of the innovation (fixing personnel to routes) from the umbrella innovation (Japan) caused this. Backfiring at the project level did not impact the innovation on the program level.

Regarding the start of the implementation of the project DSSU, based on this Japan idea, ambiguity helped in diminishing initial resistance to the plans. The innovation is systemic and will affect many other technical and social parts of the system later on, but the ambiguity helps to cover some of these ripple effects. Organizational entities whose cooperation was necessary then have fewer opportunities to find what the innovation actually will bring about and hence fewer chances to find reasons for resistance. In addition, given the fact that an ambiguous innovation can be perceived in many different ways, the innovation could become seen as a win-win situation by many parties, even though later on this win-win situation was unattainable. Summarizing, the ambiguity ensured that the sector was not immediately confronted with the overwhelming complexity of implementing the innovation and made sure that there was time for the innovation to serendipitously find an entrance into the regime system. As we will point out later, and already have shown in the previous chapter, the implementation of Japanese principles relied heavily on a window of opportunity and therefore needed to lay dormant until such window presented itself.

Ambiguity also seems to have a disruptive side. Respondents noted that the impact of the actual innovation implemented by the transport planning department and the project team was very ambiguous as well. Firstly, they could not foresee all the changes to affected parts, processes and people the innovation forced at later stages of its implementation. And if changes were detected, affected departments could not see at what moment these changes became significant and warranted action. The project team for instance more and more neglected the design guidelines and these deviations were individually communicated to the safety department. They had problems detecting when to take action, as the increasing amount of deviations created problems. A few deviations are common for any project, but in this case it became 40 and they felt the need to escalate. Secondly, the ambiguity endangers concerted action when the functionality of the innovation is perceived differently. If parties agree on the design, but disagree on the functionality, they will act differently when the design has to be changed to accommodate new information or a changed context. Thirdly, the perception of the innovation itself may change even when it is already being implemented. After heavy disruptions around Utrecht Central Station in 2009 and 2010, the innovation became from being a coherent set of measures all about capacity and robustness, to an innovation all about removing error-prone railway switches. This then evoked again an entire new debate about if this innovation would be valuable at all. The removal of switches, purely seen as an isolated measure, is highly impactful on the flexibility and the operational processes of NS and hence created much more resistance to the innovation from that organization then earlier on. So while its ambiguity can allow an innovation to persist for longer periods of time and enable the incumbent regime to cherry pick elements of it as they see fit, it also creates a dynamic in later stages of the process. Far less as is the case for more usual innovation processes, the ambiguity hinders in creating consensus and momentum during implementation. Figures 7.7 and 7.8 show this process.

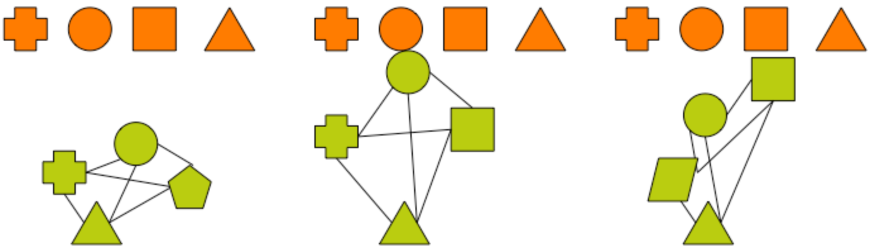


Fig 7.7. Each department can try out a specific part of an ambiguous innovation

In Figure 7.7 we see an ambiguous that can take on many forms depending on who is perceiving it. It is also malleable and adaptable. The innovation idea is able to persist for long periods of time even when it is not or only partially implemented. Even elements of the idea that backfire (such as assigning train drivers to fixed routes and the resulting controversies around 2001) do not pose a threat to the idea itself. One organizational entity (orange circle) can try out a part of the innovation (green circle) but decide to cancel its implementation without endangering the viability of the entire ensemble of innovation elements

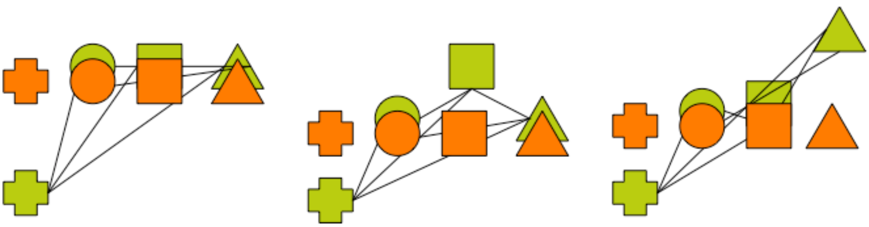


Fig 7.8. When a constellation of organizational entities does not fully grasp the inherent interdependencies between innovation elements, dynamics occur even during implementation

As an ambiguous innovation, shown in Figure 7.8, passes the organizational filter made up of different organizational entities (orange) and moves into the implementation stage, its adaptability and multi-interpretability create dynamics even during implementation. Even after moving into the implementation phase departments may neglect the inherent links between the innovation elements, as no single department fully understands the innovation

7.3.7.2 Mechanism 2: disentanglement

The earlier study visits showed that Japan was different on many dimensions of the railway system. The way stations were designed, the way timetables were planned, the way traffic was controlled all differed with approaches used in the Netherlands and all could potentially contribute to Japan's edge. This resulted in ambiguity, as mentioned above, but also resulted in little knowledge about what to actually do if one is intending to divert the evolution of the Dutch system towards Japan's path. This has to do with two things: firstly, in a fragmented sector the multi-facet nature of 'Japan' results in coordination issues. If the innovation's scope is so large, then what should we implement first and who should do it? Secondly, and more importantly, when differences between the two systems are still described very abstractly, and the lessons learned have yet to become more fine-grained, the innovation remains unworkable. That is, the innovation needs to be operationalized. Regarding the latter we see a process of operationalizing the innovation between the years of 2003 and 2010. Whereas senior managers and board members mostly undertook early visits, later visits became more in-depth and involved managers with more technical and operational knowledge. Initially Japan was seen as different on so many ways, but later visits more and more focused on infrastructural and timetable elements. Another lesson from Japan, also acknowledged in earlier visits, is the integrality with which additional infrastructure is designed and the strong feedback loop from the operational echelons. The driver of the innovation in the Dutch sector then founded the Performance Analysis Bureau to study more in-depth the actual operations of the Dutch network. He noted that this circumvents the lack of evaluation in daily operations by having a staff department studying it, but it did provide many relevant insights in what actually the problem was in the Netherlands. For instance, it appeared that malfunctioning assets such as tracks and trains was not the main cause of small delays in the system but rather that other delays triggered additional delays. Looking again at Japan, they noticed that different methods for timetabling and for traffic control were highly important. Additionally, this triggered the actors to uncover headway times as a crucial design parameter. Hence we see that more and more Japan as an idea becomes concrete but also does away much of its initial scope. In other words, the spaghetti of interrelated but highly different measures becomes disentangled. This has a constructive side as it allows innovation actors to better compare the innovation and the current system and to operationalize the innovation. However, we note that there seems to be a highly circular relation in studying both the current system and the innovation at the same time. We feel that this explains how the initial scope of the innovation became more confined to corridor separation, timetabling and headway times.

Disentangling the innovation also serves another purpose. Respondents noted that in itself the innovation is hard to bring about for two main reasons. Firstly, all the changes are costly as they involve changing the technical setup of a capital-intensive system. Next to being costly, they in themselves can provide only little impact if they are implemented separately, or coherently at less crucial nodes of the network. One respondent noted that there is a tendency for the sector to implement changes through large programs as these overcome the scarcity of attention that is apparent. Larger programs seem to be better able to persist over time and are less prone to cancellation. Other than at Utrecht central station there were other instances were actors tried to implement elements from Japan. The Top-50 project emanating from the Robuust Spoor program is an example. On 50 less crucial parts of the system, Japanese principles were used to redesign the infrastructure. However, only few of them were implemented, as the benefits could not outweigh local counter pressures.

The disentangled innovation made sure that the innovation could connect with a problem owner and with a window-of-opportunity. The problem owner was the transport planning department that was able to use the disentangled innovation. This department sought ways to more efficiently redesign infrastructure and that part of the innovation both enabled that and fell to large extent within the mandate of that specific department: these principles were mainly design and timetabling principles (their responsibility) and not construction principles (which would be the responsibility of the project department of ProRail). In addition, the department at that time had specific the requirements for the redesign of Utrecht Central station, the most crucial node in the network. Applying the Japanese principles there could immediately proof to be valuable as well as provide a turning point for the evolution of the system. Hence, the already planned rebuilding of the station functioned as a window-of-opportunity for the innovation. An innovation that otherwise could not be implemented due to the inertia of the system.

Disentangling an innovation makes sure that at some point the innovation is able to go through the filter that is a fragmented sector. It also allows for that part of the innovation to make use of windows-of-opportunity provided by already planned projects. However, in essence the innovation remains a systemic innovation and some parts of this innovation will trigger other innovations later on. It appeared especially that the part of the innovation that went through the filter and ended up attached to the Utrecht project caused a tipping point. Whereas earlier lessons of Japan involved preventive maintenance and station development and had little ripple effects, changing the infrastructure layout had many expected and unexpected implications for other processes. For instance, disentangling the innovation resulted in the infrastructure part being taken up by the focal department but as well resulted in the safety signaling part being taken up by another one. However both parts of the innovation are highly interrelated and this created the phenomenon that the project team was designing the infrastructure and more and more had to use elements of this other innovation as well. So at the moment when the innovation is disentangled and at least in perception has lost much of its systemicity, in later stages this systemicity returns. It appear that the implementation of what is left over after fully disentangling an innovation mandates the implementation of other changes, part of the innovation before disentanglement, as well. This creates the phenomenon that different departments are tackling the same innovation simultaneously, as we saw in the timeline of events. We show this phenomenon in Figure 7.9.

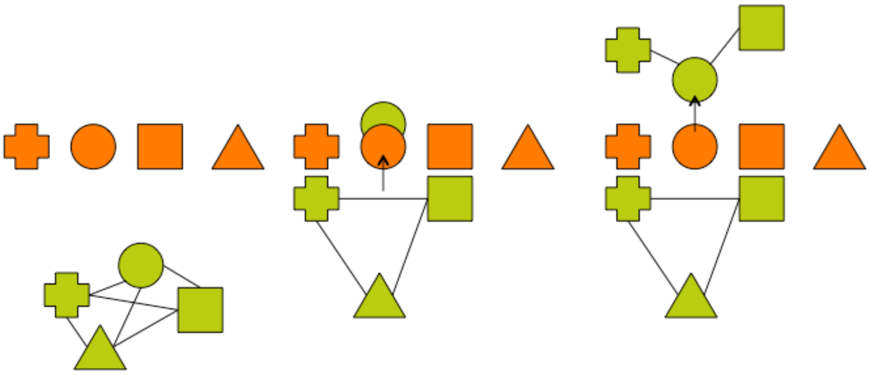


Fig 7.9 How one organization entity studies the innovation and disentangles it into a manageable part

In Figure 7.9, for a fragmented sector (orange) to pick up a systemic innovation (green), it needs to disentangle manageable parts of it from the larger whole. However, after implementation its systemicity returns as the implemented part forces other parts of the initial innovation to be implemented as well. This causes departments studying their respective part of the innovation (for instance the orange square studying the green square) while other departments have already implemented that part of the innovation (the orange circle has implemented the green square)

7.3.7.3 Mechanism 4: interlocking

The third mechanism that emerges from the data is that of continued interlocking between projects over the course of a systemic innovation process. We define interlocking as the phenomenon of two or more projects becoming structurally (sharing the same design parameters) tied or functionally (sharing the same functionalities) tied on a technical level.

In the previous mechanism we have described the benefits of disentangling an innovation as this allows for parts of the innovation to interlock with other innovations. First and foremost this has to do with an innovation finding a window-of-opportunity. In this specific case it was the already planned rebuilding of Utrecht Central station that provided a specific opportunity for a part of the Japan idea to get foothold in the current system. In a sense this interlocking between the Japan-movement and the Utrecht project provided an opportunity for a systemic innovation to find a seed of change, irrespective of potential positive or negative outcomes later on. Additionally the continuous interlocking creates ever-larger projects, which in itself ensures the already mentioned scarcity of attention. The combined top 50 project, 50 separate interlocked projects, was exactly about achieving this, although eventually only few of the 50 projects really were implemented. For Utrecht, the interlocking between the innovation and the large project ensured that it gained much momentum and could overcome to some extent a lacking support or organizational opportunism.

Besides finding windows-of-opportunity and building momentum, interlocking also occurs because of the coping strategies of actors. As we have noted there is a tendency to use assumptions as means to deal with complexity whilst ensuring progress in the project. Often these assumptions are based on designs or decisions happening elsewhere in the organization and for totally different projects. This interlocking then has a constructive side as it allows stakeholders within a project to deal with the complexity as it provides the opportunity to use assumptions rather than time-consuming additional research or other coping strategies.

Interlocking has a disruptive side as well. Firstly, when an implementation process relies on continued interlocking with other projects, its path becomes very serendipitous and hence less controllable. For the larger sector what happened in Utrecht came as a surprise. A shared and detailed conception of what Japan is and what it could do for the Dutch system was not yet fully developed, or even non-existent, but still parts of the innovation got implemented. This interlocking then deviates strongly for more linear conceptions of innovation where shared decision making in different stages of the process is prescribed or at least assumed. This shared decision-making becomes however problematic when the process is highly stipulated by outside events such as other projects allowing for windows-of-opportunity. Above that, and even more problematic, is that the use of assumptions is an easy way for innovation stakeholders to deal with complexity whilst ensuring progress, but that this merely results in technical interlocking. This is problematic for two apparent reasons. Firstly, whereas project-internal complexity is temporarily reduced, the overall network of interlocked projects becomes highly complex. Dynamics in different projects constantly impact each other. Moreover, these dynamics can be qualitative different as the data shows. Utrecht, of which at a specific stage was mostly a tactical design project, interlocked with the large freight train bridge project in the East of the country. Such a project, at that specific time, was much more a political project were technical rationale for the decision-making was absent. Rather, political support, financial support and the support of the impacted community played a much larger role in explaining the dynamics of that process. Also assumptions may rest on projects in a completely different stage of their process and hence in a different stage of built-up momentum. The design of Utrecht is fully based on a specific future timetable. At that moment, in 2010, the format of the available timetable was a basic hourly pattern of trains entering and leaving the station. With this information, the design team designed an infrastructure layout for Utrecht, disregarding the fact that the timetable would be subject to changes as well as subject to more detailed design in the later stage of timetabling. Based on the hourly pattern the design team came up with an infrastructural layout that tightly fitted the then available timetable. When the logistics department and the asset management department started to look in more depth to the infrastructure and how it would cope with other timetables, problems emerged. It appeared that thinned down timetables as well as usual shunting procedures during start and end of the daily cycles became highly problematic. The assumption that basing the infrastructure on the most ambitious timetable would render it workable for all other timetables was invalidated. For instance, under a frequency of 6 regional trains per hour to and from Houten and to and from Woerden, these two connections could be combined to create one large route with Utrecht as a throughput station. Given that a throughput station does need little switches to allow trains to shunt and change their direction, the designed version of Utrecht did not allow for a timetable where only 4 trains could go through to Woerden and 2 trains coming from Houten had to be send back. When during the process it appeared that the needed bridge extension in the west of Utrecht would be finished by 2018, this posed a large problem. So we see here that the project interlocked on a technical level with a not yet fully designed timetable and with a yet to be build bridge extension but did little to incorporate the dynamics in these other two projects later on. In Figure 7.10 we provide an abstract account of such an interlocking process.

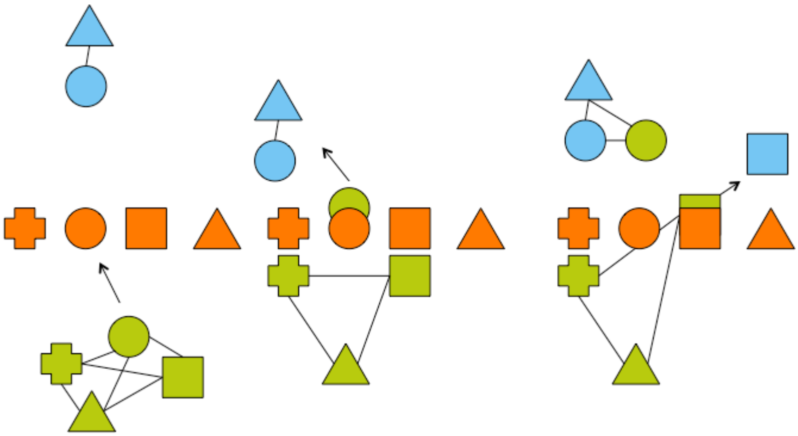


Fig 7.10 Part of an innovation interlocks with another innovation

On its own, a systemic innovation as portrayed in Figure 7.10 (green) is not valuable. It needs to interlock with ongoing change projects (blue) that provide a window-of-opportunity. As different parts of the innovation find different windows, coherence is lost in implementing the entirety of the innovation (In the beginning belonging together, the green circle and square are separated and linked to a blue triangle and circle and a blue square respectively).

7.3.7.4 Mechanism 4: shielding

Shielding concerns the extent to which an organizational entity forgoes on communicating any of its decisions or designs to other organizational parties at a specific moment in time. It also concerns how communication from outside is blocked, communication in the form of other decisions, norms and rules or demands. A natural level of shielding exists in the sector and as we have shown in the static context factors, this resolves around the high professional fragmentation in the sector. Under normal circumstances, professional fragments are shielded from each other and they enjoy relative autonomy in dealing with their respective environmental pressures. The coping behavior of actors only adds to this shielding over time.

Shielding proved to be effective in planting the seed for change. It allowed the department involved in bringing about the innovation to enlarge their design space to find a critical subset of the innovation that would function as a tipping point. Respondents both internal to the project as well as external noted the strategies that this department adopted relied heavily on shielding them temporarily from outside pressures. Usual ways of acting, stipulated heavily by norms, rules and institutions, were discarded intentionally by the transport planning department to allow them to design Utrecht from the ground up. In addition, they actively increased their design space by not only designing infrastructure deemed valuable by their customers (train operating companies) but they also designed a timetable in advance to more tightly interlock the design of the infrastructure to some hypothetical timetable. Hence the department was designing far beyond their usual design space boundaries. This shielding therefore adds to the constructive sides of ambiguity, disentanglement and interlocking as it allows that part of an innovation that gets a foothold to be large enough to still be a tipping-point to the system and not a mere incremental change.

During the process shielding also allows for more easy interlocking processes. This has to due with the fact that without organizational boundaries having to be overcome, using assumptions becomes much easier. For the stakeholders involved in the project it allows them to constantly interlock with other projects without also instigating effective (but time consuming) coordination mechanisms between these projects. The interlocking mentioned in the previous mechanisms, such as with the freight bridge, the bridge extension and the timetable, could only contribute to progress if those usually responsible for these projects are not incorporated into the direct stakeholder arena.

The downside of this mechanism is that as we have shown interlocked projects tend to be of a different kind, in different stages of development and subject to different dynamics. With shielding and the accompanied interlocking patterns, we see that projects tend to become interdependent on a technical level without having the stakeholders around it becoming interdependent as well.

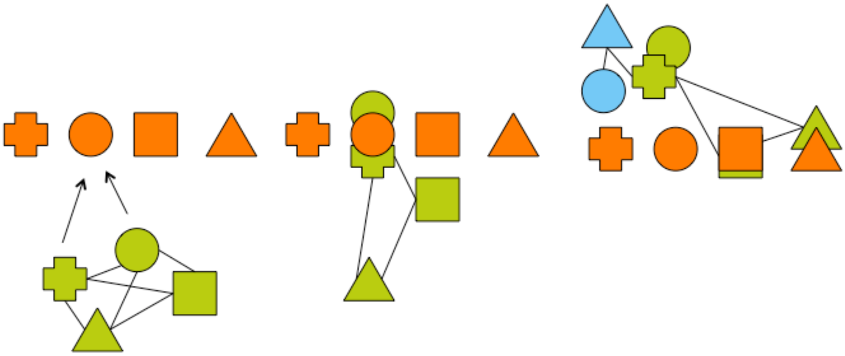


Fig 7.11 Shielding to increase the design space of one organizational entity (the 'orange circle' department implements a green circle and a green cross as well)

Figure 7.11 shows how shielding increases the design space. Here, shielding enables fragmented organizational entities (orange) to pick up more of the innovation than they are solely responsible for, hence one single organizational entity increases their design space. By doing so, they can attach more of the innovation (green) to ongoing projects (blue) and are better able to provide a tipping point to the system as well as make sure that systemicity of the implemented innovation is maintained by having interlocking over a range of other projects at a minimum. In this case one organizational entity (orange circle) takes into its design of the innovation an element for which another department would actually be responsible (orange cross)

7.3.8 Interaction effects

Although we have discussed the four dynamic mechanisms separately, they are continuously interacting with each other. This interaction is partly the reason why constructive mechanisms become disruptive, because mechanisms moderate the beneficial influences of other mechanisms.

For instance, shielding is said to be persistent over time. Building niches, either internally constructed by actors or externally produced by structural properties of the sector, is therefore not something that is always beneficial to an innovation process. The persistence of shielding, even when new actors enter the innovation system and old actors leave, is caused by the moderating effect of other mechanisms. The project team that had to design DSSU based on the specifications of the transport planning department was faced with so much complexity that it had to increase shielding again. The transfer of complexity from the transport planning department could only happen because they kept many of the details ambiguous and operationalized the innovation solely to what they as a department were responsible for. Furthermore they interlocked their innovation to the Utrecht project, a project that was not solely about introducing Japanese principles but also about other improvements. The transport planning department was only able to do this by shielding themselves temporarily from the pressures of their outside environment, but given the mechanisms that this shielding enabled, the shielding mechanism lingered on. Even when the transport planning department was no longer the lead in the project, the new lead stakeholders (the project team) had to continue shielding their design arena to cope with the complexity.

The disruptive side of shielding also has to do with several innovation elements being interlocked with separate projects. So whereas interlocking is effective to get a systemic innovation going, the moderating effect of shielding, a mechanism that is persisting because the separate projects are faced with increased complexity, turns the effectiveness of interlocking in something disruptive. The result is that the original elements of a systemic innovation are dispersed over a set of individual projects and that coherence between these projects is lost. So we see that signal optimization is becoming both a separate project in itself as well as part of the DSSU project. We see that switch removal is becoming both part of DSSU as well as part of other projects trying to reap the benefits the sole removal of switches cause. Disentanglement also plays a role here because the lead department driving the innovation before the design stage (the transport department) more and more focused on only the elements they could really control given their role in the process, creating less coherence in the initially envisioned transition.

Ambiguity is effective to keep an innovation floating as an idea, allowing the idea to wait for disentanglement and interlocking with ongoing projects. Basically the latter two processes can be done because ambiguity is effective in instigating the mechanism of shielding. Because the full details of the innovation idea were not fully known, the involved stakeholders created shields from immediate pressures from those potentially opposed to the idea. Shielding and disentanglement led the project team to base their decisions on many assumptions, a pattern we also saw leading to a continued shielding over time. Given these mechanisms, the innovation remained ambiguous, at least to those not directly involved in the project but still heavily impacted by the innovation’s repercussions.

7.4 Synthesis

We have depicted four mechanisms that describe in more detail the processes that are instigated when a systemic innovation is triggered in an inert system. All four mechanisms had in them constructive and disruptive components and we have seen that over time a mechanism switches from being the one to the other. In this paragraph we distill one general pattern that describe the trade-off apparent in all four patterns. This general pattern is highly related to the macro-level pattern we found in the previous chapter.

7.4.1 Chaotic exploration

The constructive sides of the four mechanisms deal mostly with allowing seeds of change to occur in a system otherwise hostile towards all too radical change. We briefly summarize this:

- A radical innovation is able to persist over longer periods of time parallel to the existing system
- Immediate resistance to the innovation is overcome
- The exact working of the innovation can be experimented with and is hence more operationalized
- The innovation is able to find a window-of-opportunity in the existing system
- Innovation actors are able to decrease the technical, social and institutional complexity inherent to bringing about radical change. By shielding their innovation project from outside pressures they can use assumptions as a way to effectively decrease complexity and ensure progress.

7.4.2 Orderly exploitation

The disruptive side deals with coherence that is lost when the mechanisms persist during the innovation process. Ambiguous disentangled systemic innovations that interlock with other processes of change whilst the innovation actors remain shielded from their environment create many disruptive phenomena. Most of these phenomena encompass the inability of the entire system to deal with dilemmas in a coherent manner. These phenomena are the mirror image of phenomena deemed constructive in providing the needed seeds for change. As innovation actors cope with the complexity of bringing about systemic change they seem to push dilemmas away from them in space and time. Other departments at the same time or later in the process are faced with the dilemmas but find little opportunity to deal with them accordingly. Summarized we see the following:

- Dilemmas are not recognized as separated organizational entities deal with only their side of the dilemma.
- Dilemmas involve values of different kinds of measurability. Measurable values seem to be prioritized
- The sectors institutional setup is not designed to deal with these types of dilemmas when they are recognized.
- Opportunities to deal with dilemmas present themselves far after the dilemmas have emerged; all the while ongoing progress pushes for one specific side of the dilemma.

7.5 Theoretical explanation

In Chapter 3 we have elaborated on a range of theoretical frameworks dealing with systemic change. Three frameworks seemed most relevant: the Multi-Level Perspective, Technological Innovation Systems and the NK-model. These frameworks helped in disciplining our analysis of the raw data. However, in doing so we were also able to uncover where the raw data seemed to point to gaps in existing theoretical frameworks. In addition, the mechanisms that we distilled from this raw data do not fully overlap with the concepts, phenomena and dynamic patterns propagated by these frameworks. In this part of the chapter we try to make sense of our findings using these frameworks and see where current models fall short in explaining the mechanisms we found.

7.5.1 NK-model

As already mentioned, the NK-model is a highly abstract and formalized way of portraying the evolution of complex systems over a fitness landscape. Much of our mechanisms emerge just when we abandon this abstraction and introduce the messiness of systemic change in a real-life setting. Organizations, institutions, actor strategies, politics all play a role but have been purposefully neglected in NK-modeling. Notwithstanding these differences, NK model still provides useful insights for making sense of our findings. The model provides a nomenclature as well as abstract dynamics when we speak about and analyze a system that changes from one state to another.

Regarding the constructive side of ambiguity, the notion of fitness landscapes from NK-models is important. Fitness landscapes are topological representations of the optimality of all system configurations. A systemic change is move from one configuration to another, which does not belong to the same basin of attraction. This inherently involves traversing a valley, steps in between the current state of the system and the end-state that are less optimal compared to both. In our case such a valley involved both a partially implemented innovation as well as the pervasive effects of implementing the innovation on the daily operation of the system. The ambiguity of the innovation here caused two things: firstly, fitness landscapes are multi-dimensional and what is considered ‘fit’ by one actor might be considered ‘unfit’ by another. A certain ambiguity made sure that all actors perceived, to some extent, the end-state as an improvement over the current state. Secondly, the ambiguity of the innovation made it hard for the involved actors to analyze the exact valley that would have to be traversed. This valley only became apparent after the end-state was defined as desirable by a large enough coalition and the process already built up some momentum.

NK-models also acknowledge that jumps from one optimal point to another involve multiple steps. However the model disregards the fact that for different steps, different actors are involved. Also, these different actors perceive the fitness landscape differently. A jump from one optimal point to another in our case also involved disentangling the innovation. This created two phenomena in NK-model terms. Firstly, continued studying of both the Japanese and Dutch system caused a more detailed picture of the exact location on a fitness landscape of both systems. This exactness was however traded in with comprehensiveness as the search for knowledge more and more focused on parts of the innovation. Due to this pattern, the innovation champion had more knowledge on what specific step had to be taken to arrive at a new basin of attraction while losing the more optimal point out of sight. Also by doing so, it became more unpredictable what other changes would be instigated once this innovation was implemented. In NK-model language, the innovation traversed a valley and but the resulting topology of the fitness landscape, given other actors actions, was still unknown. The innovation may be very well get stuck in a local optimal point yet again.

7.5.2 Technological Innovation Systems (TIS)

As we have stated in our theoretical chapter, TIS mostly focuses on the internal functioning of an innovation system. However, a structural component, that which brings about these functions, is underappreciated. In Chapter 6 we have delved deeper in these structural components using the PSI framework and also in the analysis in this chapter the structural components played a prominent role.

Adding this structural component brings to light the most important contribution of this analysis. In the TIS literature the focus is on reinforcing functions, functions of an innovation system that cause other functions to arise as well. The analysis of the structural configuration of an innovation system underneath these functions however shows that rather than reinforcing functions, functions can also work counterproductive. Hence, optimizing innovation systems is not merely looking at which functions are not performed and adapting the innovation system accordingly, but much more a careful weighing of important functions at a specific moment in time. It appears that the relations between structural parameters and functional parameters of an innovation system are highly polygenic and pleiotropic, meaning that optimizing on functions is a complex task. For instance, while ambiguity is effective in diminishing resistance and strengthening the network of actors, knowledge development activities, a second function in the TIS framework, counters this. The two functions 'knowledge development' and 'network building' thus can interfere with each other through their shared link with the mechanism of innovation ambiguity. In other words, sometimes a lack of knowledge is effective in overcoming early resistance to the innovation.

Disentangling an innovation does away with much of the ambiguity, but if done in a shielded environment the effects are much less. However, as we have seen, more in depth knowledge on the innovation is negatively related with a broad comprehension of it. Here we see, as TIS shows, that learning activities and market formation are correlated functions, although we have to stretch the reach of the TIS framework a bit. As the innovation actors were learning about the innovation, they disentangled it to such an extent that they were able to interlock it with ongoing projects. In terms of the TIS framework, in a hierarchical environment they were able to form a market for their solution: a parallel project that sought the very benefits the innovation was providing. However in contrast to TIS, we see that this seriously impacts the technical makeup of the innovation, from a broad collection of changes to more narrowly defined changes. Again this structural parameter influences functions of the innovation system later on. For instance we see that multiple innovation systems arise around the innovation idea, each system involved in studying and implementing their respective part of the innovation. The fact that out of one innovation system multiple innovation systems can arise is an underexplored notion in the TIS literature, and possible interesting to study.

Interlocking is the main pattern that the general TIS framework is poorly able to explain. TIS has been generally seen as internally focused, looking solely at the internal workings of technology, innovation actors and institutions (Markard, 2012). Context, which in our analysis seems to have a pervasive effect, has only gotten little attention. Lately however, given calls to combine TIS with more contextual frameworks, such as MLP (Markard, 2012) more work is done on introducing context to the analysis of innovation systems. As of yet, a mere typology of different context-innovation systems relations is present. In our analysis this interlocking occurs out of two reasons: an internal drive from the innovation system to deal with complexity and an external pressure from the context (i.e. the regime and other innovation systems) to reap the benefits from the focal innovation system as they see fit. We have seen that interlocking first of all allows change to occur in the first place but that this comes to detriment of heightened dynamics at later stages as multiple projects become technically interdependent. This 'hyper-volatility' is something worth introducing in the TIS realm, as the effects of contextual links have yet to be studied by the TIS community

In essence TIS deals with shielding as it portrays innovation systems in isolation. If no contextual influences impact the dynamics of a TIS, the innovation system is perfectly shielded. We have seen how acting like an isolated TIS gave many degrees of freedom to search further on fitness landscapes, allowed for time to disentangle an innovation without decreasing ambiguity, and provided opportunities for interlocking with other project without having related actors and institutions joining the innovation system. The key notion here is that as actors join a TIS they bring with them new perspectives, incentives, strategies and institutions. This provides chances for growth, as the TIS framework posits, but also provides constraints. Shielding is a pattern that gets its constructiveness from this very effect, by alleviating usual constraints due to the entry of new actors.

7.5.3 Multi-Level Perspective (MLP)

Of the three frameworks, the MLP is probably the most valuable in making sense of our case study. The notions of niches, regimes and momentum-building also played a role in analyzing the raw data. This dual usage of the framework results in both seeing how helpful the MLP was in making sense of rich qualitative data and in what ways the resulting conclusions regarding the mechanisms overlap with the MLP. The MLP deals with systemic change in path dependent sociotechnical system and the continuous interaction between a hostile regime and an innovation in a niche. Much of the features explicitly mentioned in the MLP literature can be found in our analysis. Although these features relate directly to the four found mechanisms, they mostly deal with one side of these mechanisms. Hence, if we apply the MLP framework we miss the dilemmatic nature of these processes. This is due to the MLP's focus on competition between niche innovation and regime, rather than a cooperative effort. Whereas in a competitive environment tradeoffs do not play a role as each side of the tradeoff belongs to either the innovation or the regime, in a cooperative effort a choice needs to be made by the entire arena of stakeholders from both the niche and the regime.

A good example can be provided by our first mechanism about ambiguity. In later stages of the process it appeared that ambiguity lost its constructiveness. A lack of a shared vision meant that actions were not aligned and that new entrants to the innovation arena heavily influenced the technical makeup of the innovation. In the MLP literature a high focus is given to vision setting and experimenting, finding increasingly the exact structure and functionality of the innovation. Together with network-building, these three forces impact each other and give the process a certain momentum. This momentum is needed since over time the niche innovation more and more interacts with a hostile regime. As we have found, ambiguity, related to a lack of vision building and learning, did cause the innovation process to lack momentum. However the other side of the coin is that ambiguity also had its constructive side. In the MLP framework this notion is less prominent. The framework mostly deals with a given innovation that has to be implemented by 'change-inclined regime players'. However given that in our case the regime had to pick up the innovation, a window-of-opportunity was needed. This is crucial to our finding that an innovation needs to have some form of longevity even when it is not being implemented. The process is very temporal and serendipitous and hence demands from an innovation to lay dormant to increase the chance that the 'idea' is still there when an opportunity presents itself. This ability to lie dormant is directly related to its ambiguity as regime players can perceive the innovation as they see fit and little to no dilemmas immediately are conceived during the earlier stages of implementation. In MLP terms this would involve two things: a lack of vision building might be beneficial if a niche innovation depends on yet unknown windows-of-opportunity. Secondly, a less confined and more ambiguous vision might help in overcoming resistance and help in network building. This again then places constraints on experimenting, as this impacts the specificity of the vision.

Experimenting in niches is a focal point in the MLP framework. We have seen the same in our case study. In a more fragmented sector after the split-up, parts of the ProRail organization were able to discover and analyze Japan as a suitable benchmark and they were able to more freely design infrastructural plans without the immediate institutional pressures of the regime system. Albeit that this learning did not so much occur through experimentation, but rather through studying a system as a benchmark, having niche-like qualities supported the learning process. However, contrary to usual accounts on processes using the MLP, we see that learning qualitatively impacted the innovation itself, this impact we call disentangling. In the MLP framework, albeit not mentioned explicitly, the general notion on learning about innovation is that the innovation grows quantitatively rather than changes qualitatively. Again, this has to do with the big role the regime played in our case study. The regime functioned as a filter system, and only allowed certain parts of the innovation to go through and interlock with ongoing projects. While disentangling helped in dealing with the filter function of a regime characterized by a fragmented organizational setting, it also caused the innovation to lose partly its systemicity, at least midway through the process. It regained much of its systemicity, as other changes were forcefully instigated due to the tipping-point qualities of the residual changes (i.e. those changes that survived the 'filter'). This technical dynamic of the innovation artifact is highly important in explaining how and why the process came to be, and cannot be fully explained using the MLP framework.

Interlocking has been part of the MLP literature and therefore the framework provides opportunities for making sense of our case. In the MLP literature interlocking mostly deals with the innovations' ability to create momentum by having an ecosystem of mutually reinforcing innovations 'attacking' the regime. Interlocking can create windows-of-opportunity and we have seen in our case that the rebuilding of Utrecht provided such an opportunity for the innovation. However, contrary to the more convergent interlocking that usually is seen in MLP frameworks, more divergent interlocking is also possible, where more and more elements of one innovation spread out over different projects. In the timeline of events we showed how the original idea of Japan got spread out over multiple projects, each focused on implementing a subset of this idea. This is highly interesting since it involves the take-up of systemic innovations by regimes whilst at the same time the innovation loses its systemicity. Elements of systemic innovation only provide full benefits if they are implemented coherently, but the very fact that the elements have to 'survive' on their own in projects started for different purposes than what the elements were meant for is an interesting dynamic. The extent to which the elements, once implemented, still are able to complement each other and provide the synergetic benefits inherent to systemic innovations is doubtful.

The core of MLP deals with niches and shielding is just that. However, the MLP assumes niches to disappear once niche-like qualities of innovation systems are no longer deemed necessary. Our analysis however shows that niches tend to persist. For dealing with complexity niches are highly effective. In a shielded environment, innovation stakeholders can take time, use many assumptions, and disentangle the innovation to such a point where they see opportunities for interlocking with ongoing projects. The resulting complexity for later stakeholders creates the persistence effect of the shielding mechanism. This persistence of the shielding mechanism creates the effect that the innovation and the regime no longer constructively interact and the elements of the innovation, now interlocked with other projects, no longer are perceived in their coherence to each other. Especially for the normative management theories on niche-building, highly influenced by the MLP, this is an interesting finding. The strategic niche management literature could incorporate this notion to more fully understand the costs and benefits of building niche environments. This is not solely about niches creating infeasible unrealistic utopian innovations that have no place in real-life, although this is inherently partly a problem, it is more about niches having a self-reinforcing effect and that building niches is just as much about breaking down niches in a controlled fashion later on. By whom, when and how is interesting topic for further research in the strategic niche management realms. Table 7.3 presents an overview of the relation between the MLP, TIS and the four mechanisms found in this chapter.

Table 7.3. The four mechanisms in relation to the MLP and TIS frameworks.

	MLP	TIS
Ambiguity	MLP focuses on the niche accumulation but neglects how a niche can also survive if accumulation does not take place. The portrayal of transitions starts as a seed grows to finally takeover a system. Rather this analysis shows that transitions are serendipitous and the ambiguity of an innovation helps in connecting an innovation with a window of opportunity	TIS is centered on the innovation system to the detriment of a focus on regime and innovation interactions. Ambiguity helps in making this interaction less confrontational. Also TIS postulates that core functions of a TIS are needed for bringing about systemic change and that these functions can enforce each other. Lack of immediate experimentation (ineffective according to TIS), which increases ambiguity, is however seemingly beneficial to some extent.
Dis-entanglement	MLP postulates that as a niche innovation grows, it does not categorically change. However, for a change to occur in this case, the innovation had to change from an entangled set of many interrelated measures towards a more operational subset of it.	TIS barely deals with the existing makeup of the innovation as it is more focused on the functions of the entire system (including actors and institutions). However the structural qualities of an innovation appear to matter and also impact how actors and institutions evolve over time.
Interlocking	MLP mainly describes the individual journey of an innovation from niche to regime, albeit interlocking has got some attention. Interlocking is merely discussed taxonomically, this analysis shows what happens during interlocking MLP deems interlocking effective, this analysis shows that it can also be disruptive	Synergies or conflicts between different innovations falls beyond the scope of traditional TIS studies
Shielding	Niches are deemed to open up over time, this is not self evident given our case study: niches are persistent and actors inside the niche actively cause this persistence.	TIS portrays the opening up of niches as automatic, given the reinforcing effects of the fulfillment of different functions. Our case study shows that functions can contradict and that what opens up a niche, can also limit the amount of knowledge development. That is why the growth of technological innovation systems is not automatic.

7.6 Mechanisms and patterns

In the previous chapter we found that systemic innovation processes are distinct from other innovation processes in that there seems to be less built-up of momentum and that volatility in the product, social and institutional spaces is located at the end of the process. This contradicted usual notions on innovation processes, notions both stemming from linear and non-linear perspectives on these processes. This macro-level pattern was uncovered by solely looking at the project-level or innovation system level dealing with implementing the innovation. By looking more broadly in space (beyond the mere project) and in time (tracing the origin of the innovation from its initial discovery), we were able to uncover four dynamic mechanisms.

Given the four dynamic mechanisms at work during systemic processes that we have uncovered in this single case study we see why the macro-level pattern of volatility in the structural features of an innovation system is distinct from other innovation processes. The four mechanisms lead to motors for change by pushing away dilemmas from the immediately involved stakeholders. Systemic innovations are inherently complex and dilemmatic and resistance from the regime is overcome by the constructive properties of the dynamic mechanisms. On a project-level this leads to little changes in the P, S and I spaces: shielding leads to few new entrants to the innovation system; ambiguity leads to little uncovering of the true complexity of the innovation; disentanglement in conjunction with shielding leads to a P-space where the product becomes more operationalized yet more dependent on assumptions on the developments in other innovation systems. Finally, interlocking creates the effect that the initial innovation disperses over several independent innovation systems. By solely looking at DSSU and backtracking the origins (as we did in Chapter 6) we see that volatility in the initial stages is transferred to other innovation systems (a pattern we see by looking at the entire ecology of innovation systems (as we did in this chapter).

For other types of innovations these mechanisms play less of a role and hence explain why the patterns of pushing away volatility in time and place is both less prominent and less feasible. The TMS project at Network Rail involved a highly internally interdependent set of elements and needed much upfront investments. Hence keeping it ambiguous was less important and even impossible. The dilemmas involved in introducing a new traffic management system and the new roles and rules for traffic dispatchers were to some extent easily detectable by those not yet involved in the project. Furthermore dilemmas were also avoided later on by keeping the innovation system both technically and socially independent from other innovation systems by building interfaces and by restricting the design space in each innovation system. These externally determined restrictions in design spaces were a conscious effort to alleviate shielding: project members had to deal with the complexity and could not transfer it to other projects through the use of interlocking. A similar phenomenon we also saw at the tunnel project where the elements of the tunnel were separately designed, built and tested, but the interaction between these elements were predetermined and related to each other in a modular fashion. As long as project managers and designers responsible for single elements kept in mind the restrictions in the design space the entire elements would fit together at later stages of the project. Both the other cases show that the internal architecture of the innovation mattered (half a TMS is not a TMS and half a tunnel is not a tunnel) whereas the coherence was more easily lost in the DSSU case. Given that coherence was so important for the first two cases many of the dilemmas had to be resolved upfront, resulting in dynamics in the P, S and I spaces. In the DSSU case the motor for initial change was the opposite, the effective pushing away of dilemmas in space and time. Given that dilemmas in the end always pop up, especially when the innovation and the regime start to more strongly interact, dilemmas seem to emerge structurally at later stages of the process.

Finally we wish to add to the analysis in Chapter 6 that we have seen that innovations are not radical or systemic in essence but that the properties of an innovation are also the result of purposeful actions during the process. In the DSSU case we see that the initial innovation idea was highly systemic but lost much of its systemic properties once it had to go through the so-called organizational filter. However, given that actors involved in bringing it through the filter (the transport department) had disentangled in such a way that the implemented changes would force many other changes later on the innovation retrieved some of its systemicity later on. This to the surprise of stakeholders not directly involved but heavily impacted. Hence, the level of systemicity in time would be a U-shape. For the TMS case the initially planned changes were also more systemic, involving a set of changes highly related yet not structurally connected. However, project managers purposefully made it less systemic by not seeking for synergetic effects with other innovations or with developments in the regime. The building of technical interfaces with other systems is exemplar and more so was the decision to buy the TMS off-the-shelve. Hence what is eventually implemented is more a radical change than a systemic one. In Chapter 6 we have made the analogy with block add-ons (TMS) and sand-invasion (DSSU) to show the difference, but other terminologies might be more applicable.

7.7 Conclusion

The analysis of the interviews shows four general mechanism that emerge when a systemic innovation process takes place in a sociotechnical system that is fragmented and involves a path dependent and inert technical system. These mechanisms we have labeled as ambiguity, disentanglement, interlocking, and shielding and are constructive in the beginning but by their very nature become disruptive later on. We find these mechanisms relevant since they help us explain why the unique pattern of volatility, as the previous chapter showed its location at the backend of the process, exists for systemic innovation processes. Other mechanisms are at work, but become less relevant when solely looking at a systemic process progressing through time. These other mechanisms might also be present at incremental change or during simple improvement projects. In supporting specifically systemic change, it is through working on these four mechanisms and eventually alleviating disruptive volatility in the process that games can contribute. However problematic is that these mechanisms have both constructive and disruptive consequences to the process. This notion is the core of the conclusion the analysis proposes. The process of systemic change is inherently dilemmatic in itself. Fixing the disruptive side might make the constructive side less constructive, and strengthening the constructive side might make the disruptive side more disruptive. For a sector that is facing such a dilemma there is seemingly no one best way to tackle it.

Chapter 6 and this chapter together provide the innovation scholar and practitioner with a new conceptualization on how to perceive, understand and improve systemic innovation processes. We stressed that for these processes the increase in volatility in P, S and I spaces, or in more practical terms the increase in dynamics in designs, actor arenas and applicable rules and norms over time, is the central problem needed to be tackled. With the addition of the four mechanisms we show the motors driving the innovation process as well as how these mechanisms create the tendency of the fuzzy back-end of systemic innovation processes, rather than the more commonly perceived fuzzy front-end of innovation. Finally, the inherent dilemmas involved in bringing forward a systemic innovation process show that optimizing such a process is highly problematic or even infeasible. However, the fact that innovation processes are rife with dilemmas and this chapter's pinpointing what those dilemmas are, show innovation practitioners what to expect when embarking on a journey of systemic change in network-type industries. This knowledge on what to expect and how alleviating some of the foreseen problems might create problems elsewhere is valuable for two reasons. Firstly, the lessons of this chapter do not implicate that if implemented one needs an external consultant to predesign the structure that induces systemic innovation. Contrary to the MLP, which seems to focus more on structure than on agency, the lessons brought forward in this chapter empower the innovation incumbent rather than the outside observer. In the end, it is this innovation incumbent, actors actively involved in the process itself, that have local knowledge on what designs are feasible and in what ways implementation in their specific sector is realistic. Doing justice to the agency of these stakeholders and staying away from too much focus on structure has as a consequence that no outsider to the process is expected to be able to design a systemic innovation process beforehand. As we have seen at the DSSU case the process is simply too complex and too contingent upon many contextual influences that such design is infeasible.

The lesson of this chapter also implicate a new understanding on how gaming simulation can be valuable in supporting systemic innovation processes. Given the abovementioned proposition that incumbent actors themselves are best suited to improve such a process, their use of a gaming simulation to manipulate certain mechanisms to intelligently control the amount of dynamics over the course of the process might be key. So we state that, ideally, through impacting certain mechanisms, gaming simulation enables innovation stakeholders to alleviate to some extent the problematic increase in volatility in later stages. Given the complex nature by which mechanisms interact to create this pattern and the fact that these mechanisms are both constructive and disruptive, the way gaming simulations should be designed and employed becomes both relevant and complex. In the next chapter we will delve deeper in the role of gaming simulation and how this tool can alleviate volatility through impacting the four mechanisms of ambiguity, disentanglement, interlocking and shielding.

8 The Functions and Dysfunctions of Gaming Simulation for Systemic Innovation

In Chapter 3 we theoretically deduced what a gaming simulation should do to support a systemic innovation process, given two models on the relation between game design and innovation processes (see Thomke, 2001 and Klabbers, 2009). We found that merely adopting an analytical science perspective to the design of gaming simulation (the Design-in-the-small as pure experiments) will not relate constructively to the Design-in-the-large of systemic innovation processes. Chapters 6 and 7 then intended to better understand this Design-in-the-large, as potential context-of-use of gaming simulation, and uncovered leads for performance indicators via functional and dysfunctional mechanisms at work.

Both parts of the thesis, the theoretical exposition of the potential role of gaming in Chapter 3 and the empirical Chapters 6 and 7, serve as input for this chapter. This chapter will look at how these games correspond to the micro mechanisms and macro patterns from the empirical chapters. Finally, this chapter ends with more normative considerations: given the links between game design and innovation context, how should one design a game? Parts of the notions in this chapter can also be found in Van den Hoogen and Meijer (2015a; 2016)

This is relevant since the design of games has mainly been approached from two sides: the policy sciences and the educational sciences. Therefore, the conceptual frameworks by which gaming simulation is approached are heavily influenced by the respective dominant paradigms and nomenclature. With regards to games for policy, with which our gaming experiments seem to have more in common than with educational games, the works of Duke and Geurts and their 5C framework (2004) has been dominant. Games work and they work because they allow for communication, creativity, consensus, grasping complexity and commitment to action. Underlying this model is however that creativity, consensus and communication are phenomena always desirable. This may be true in the specific instances where we apply games for policy making (in times when there is for example groupthink, conflict or heavy fragmentation) but as we have demonstrated innovation processes are temporal phenomena with mechanisms that sometimes work and sometimes are counterproductive. It is also often not clear when to employ a gaming simulation since problem and symptom might be separated in time and place: volatility at the backend of the process was sometimes the result of design decisions made at the frontend. This makes it problematic for those wishing to use gaming simulation to support innovation processes. If there is problematic cooperation and a policy game can alleviate this, the interventionist effects of games are better understood. If there are patterns of shielding and interlocking that continuous interplay for years and tend to come to surface as problematic only in the final stages of a process, for the practitioner it becomes hard to determine how and when to conduct a gaming simulation for research.

This chapter makes the first steps towards answering these questions. Having identified a macro-level pattern that in some instances, or in the eyes of some beholders, can be considered pathological and added to that an understanding of the static factors and dynamic mechanisms creating this pattern, we are now in a better position to grasp the functionality of gaming and how this functionality can be achieved. Key to the previous chapters is that systemic innovation processes are rarely processes where lack of and search for knowledge is the core driver of process dynamics, and hence demands from us to refrain from solely perceiving gaming simulations as tools to increase levels of knowledge. Rather, as we have demonstrated, games should manipulate volatility somehow and we pose that doing so involves influencing the mechanisms such as interlocking and shielding. That is of course, if gaming simulation is at all able to do so.

The approach taken here is similar to the previous chapter in that it is highly inductive. The main intention is to arrive at a set of propositions regarding how gaming simulation for research impacts systemic innovation process volatility, mediated by the four mechanisms, and what game design aspects seem to play a role here. The inductive approach is a result of both practical concerns as well as methodological concerns. Firstly, we had no experimental control over the games that the sector chose to conduct. As problems emerged during organizational processes that were beyond our reach, the connection between problem owner and us as game designers was to some extent serendipitous. In addition, the limited amount of games we designed prohibits us from using more deductive approaches. Methodologically, we can argue for the inductive approach because of the lack of existing theory on games for research and design. We do not yet know where to look for in games and what aspects make them work and therefore we had to cast a wide net to achieve a large diversity in games. For theory generation we need games that greatly differ in their design, their execution and their place in specific processes. For instance, if all games had the exact same level of detail, then it would be infeasible to determine if 'level of detail' has any relation with the mechanisms from Chapter 7.

This chapter contains three sections that each deal with a specific relation between gaming and innovation processes. Firstly, we return to the three case studies mentioned in Chapter 6. For those cases we have either designed games or were closely involved in the design process (which is the case for the tunnel project). We have analyzed these cases from start to finish and this allows us to focus on games as part of longer term processes with both their own dynamic and a dynamic impacted by the use of gaming. For this purpose, the main analytical framework used is the PSI-framework. This part of the chapter then deals with how games have changed the dynamics in the P, S, and I spaces.

To study more immediate effects we use the mechanisms from Chapter 7. Because we zoom in on the immediate game design and execution process and leave out the context, we can use a broader range of games that we have designed. For instance, we were asked to design the BIJLMER game and this allowed us to also see to what extent this game attenuated or alleviated interlocking processes, although the project itself was not a topic of this thesis. However beyond these influences we were not able to see how process volatility was managed.

The third and final part of this chapter brackets games into two types: explanatory and exploratory games. Here we look at the extent to which this distinction is possible in real life projects, what the impact is of previous dynamics in P, S and I spaces and how game designers can control for this using specific game design decisions. This part is therefore the most normative part of this chapter, and next to a framework on debriefing in Chapter 9, the most normative of this thesis.

This chapter results in conclusions on the value of gaming simulation for supporting innovation processes. Given that with this chapter the thesis analyzed the role of gaming from both a pre-existing theoretical framework (Chapter 4), and from a framework derived by our own analysis (this chapter), the conclusion will also encompass the relevance of using the main case studies in studying gaming in relation to using solely theoretical frameworks such as linear models or the MLP.

8.1 Games and PSI spaces

For this part of the chapter we provide a multiple case study on projects that differed on the specific location in time of volatility in the P, S and I spaces and incorporated the use of gaming simulation as means to explore or test certain innovations. As mentioned in chapter 6, they differed along two dimensions: the internal interdependence of innovation elements and the overlap of innovation elements with external elements from concurrent innovations or the incumbent system. The three cases are:

SPOORZONE DELFT (SZD) - The building of a railway tunnel through the city center of Delft. A highly systemic product, which shares much overlap with the regime given technical standards, operator procedures, and safety installations.

TRAFFIC MANAGEMENT SYSTEM (TMS) - The radical overhaul of traffic control processes in the UK combined with a new traffic management system. A highly systemic product, which over time created much independence from the environment (concurrent innovations and the regime) through the purposeful design of interfaces.

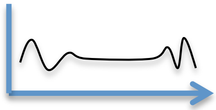
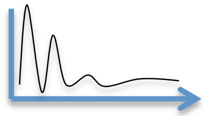
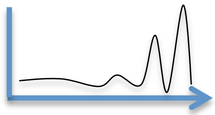
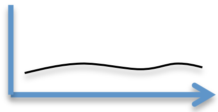
ROBUST RAILWAYS (RR) - The introduction of Japanese design and operating principles in the Dutch railway system. A collection of loosely related elements, which involved the reconfiguration of already existing regime elements. Hence, the niche and the regime, and other innovations, had much overlap

Before zooming in on the use of gaming in these cases, we provide a brief recap of the finding from Chapter 6. It appeared that the amount of volatility a certain innovation allows for and its location in time is highly determined by the rigidity and the momentum that is created by actor networks investing heavily in this innovation. When an innovation artifact is fully developed and many parties have invested considerable time, money and effort in this innovation, the entrance of a new actor in the network will cause little changes in the innovation artifact. We then found that, given this notion, certain types of innovation artifacts allow for this build up of momentum (TMS, and to a lesser extent SZD), whereas others were more suitable for interlocking processes with concurrent innovations and the regime (RR and to a lesser extent SZD). Key to this notion is that innovation artifacts are rarely atomistic products but more often a set of different innovation elements. We have seen that it is the level of interdependence between internal innovation elements and between innovation elements and the environment that impact highly where, in time, the volatility was located.

Table 8.1 provides a short overview of the three different innovation artifacts and the related volatility patterns we uncovered. Note that the classifications of internal and external interdependence say nothing about the radicalness of the innovation itself. Innovations that are internally highly connected (and hence systemic) create the need for the early build up of momentum. These innovations need the collaboration of many different stakeholders in the beginning of the process. Hence we see dynamics in the P, S and I spaces mainly at the front-end. Because of its systemic nature and the momentum it builds, it is able to withstand more of the regime pressures later on. Furthermore, the innovation artifact itself becomes more rigid over time, not allowing for interlocking mechanisms with other innovations in later stages.

The mirror image is that of innovations that are internally loosely coupled but share many overlap with external elements either belonging to the regime or to concurrent innovations. The sharing with other elements often lies in the fact that these types of innovations involve the reconfiguration of regime elements. Unlike the aforementioned innovation, these reconfigurations do not involve the adding-on of some new technology. Rather they intend to reshuffle already existing elements. An example would be the idea to increase capacity on the railway network by redesigning the total constellation of railway tracks, signaling and overhead wiring (part of RR). Here the lack of momentum and the ease by which interlocking mechanisms occur later on create volatility mainly at the back-end of the process.

Table 8.1 Volatility patterns for four different innovations.

Overlap Interdependence	External high	External low
Internal high	1. Large civil engineering upgrades (SZD) 	2. System add-ons (TMS) 
Internal low	3. Reconfiguration of regime elements (RR) 	4. Simple add-ons 

8.1.1 Spoorzone Delft (1)

Spoorzone Delft involved the building of a railway tunnel through the city center of Delft, including a new underground railway station. The interdependencies between innovation-internal elements were strong, as all elements were needed for the tunnel to function properly. Furthermore the overlap with the regime was profound in that the tunnel should allow for standard operating procedures regarding tunnel operation, traffic control and emergency protocols. Regime players from the incumbent system mandate these procedures to maintain interoperability.

For handling the project, project members adopted standard systems engineering principles to structure work processes mainly related to the technical parts of the tunnel. Firstly, this involved specifying all requirements beforehand, i.e. determining what procedures the tunnel should allow for. Secondly, this involved hierarchically structuring the design: from an overall grand design, to more detailed modules beneath it. This way, coordination was mainly realized using hierarchy where higher-level designs constrain the degrees of freedom for lower-level designs. This approach caused many perspectives and disciplines to be involved mainly in the beginning of the process. Later on, when designers could work on their specific part in modules, the work became more specialized.

Respondents acknowledged that this approach has its disadvantages. Faulty modularization creates interdependence in designs without coordination between design teams. This can cause modules to not fit properly when re-integrated in later stages of the process. It is cumbersome to design modules in such a way that after reintegration the overall artifact still adheres to the requirements specified beforehand. Modules create compartmentalization and each module might be impacted by its own eigen-dynamic as well as impacted by external pressures.

At SZD such phenomena were expected given experiences with similar projects. The project team therefore put up an extensive testing regime. This involved testing elements on the module level, as well as testing conjunctions of modules and the final overall artifact. Since only the final artifact is safe enough for operation, the conjunction between artifact and procedures is only tested in the final stage of the project. As respondents acknowledged, the highest complexity could be found at this part of the process. In previous tunnel projects, given the momentum already built up by the technical artifact, actors had to choose between costly technical changes or locally adapted procedures that endangered interoperability.

Adding to the complexity for Delft was the fact that testing the entire tunnel in its final stages could only be done in a few days rather than the usual 6 months due to spatial constraints on placing temporary infrastructure. Foreseeing potential problems, the project team instigated an additional commissioning team that encompassed both members of the project as well as members of organizational entities that would eventually use the tunnel. It mainly sought to update final users on the progression of the project. It was this commissioning team that decided to conduct a so-called integrated procedural acceptance test (IPAT).

In the spring of 2013, the team conducted several days of scenario testing where representatives from all relevant stakeholders were asked to play realistic scenarios on a scaled-down prototype version of the railway tunnel. Game players were mostly members of the commissioning team with a few additional operators invited ad hoc. The prototype version included all real life software of the tunnel as well as realistic user interfaces that would be installed in control centers. Scenarios involved for instance the managing of train traffic and tunnel operations in case of a fire at the station platform. The intention of the IPAT was to show where the technical artifact and the procedures to use it would not match. In that sense, the game was highly exploratory as the involved actors had little hypotheses beforehand on where these mismatches could be found.

The exploratory nature of the IPAT proved to be highly effective. By running a multitude of scenarios in quick fashion, and by having operators not fully immersed, the session resulted in a dialogue between engineers and users. This resulted in a list of 50 issues where technical and procedural elements of the tunnel did not match well. These issues were quite similar to issues found in tests for similar tunnel projects with the notable difference that now, due to the time still available, the technical artifact was more flexible and less costly to change. Furthermore representatives from different operational entities such as traffic control and fire brigades played the game. These representatives were operationally knowledgeable enough to validly play the game. This created the effect that the debriefing allowed for a more creative exploration of solutions as well as support the coordination between these solutions. Respondents specifically acknowledged that the improved communication between different actors in the debriefing resulted in a more careful balancing between technical and procedural solutions or combinations of these two. This was, according to them, strikingly different to real life tests where collective debriefing is impossible due to distances in space and time between stakeholders and where the role of test participant and the role of designer are separated. In addition, the installation of a commissioning team made sure that all stakeholders, also those otherwise only involved in later stages, could contribute to the design of the game and relevant scenarios and participate in the simulation itself. This improved the commitment to action afterwards. The summarized findings can be found in Table 8.2.

Table 8.2 Spoorzone Delft Game.

	Usual pattern	State before game	Game design	State after game
P	Contraction of P-space due to modularization; rising complexity and amount of languages during re-integration of modules	Simple P-space, technical artifact still flexible. All elements still modularized	High realistic game model Exploratory No dependent variable used Low immersion Game partly played by representatives from S and I-space Many scenarios	Expanded P-space. Many complexities found involving the fit between technical and procedural elements System now seen from multiple disciplines (not solely technical)
S	Volatile S-space at beginning and end of process. Midway usually separated over different disciplines with little interdependence	Expanded S-space due to installation of a commissioning team. Many different languages, perspectives and high inclusion	Participative design Open session Many game players, observers and facilitators from different disciplines	Unchanged. Commissioning team remained active until final commissioning of tunnel in February 2015
I	Clear I-space for modularizing tunnel elements. Later on unclear when different elements do not fit (during integration)	Unclear how mismatches between technical elements and procedural elements need to be resolved.	Representatives of I-space played game or were observer. Created coordination mechanism on the spot during debriefing.	Clear handling of expanded P-space through institutions agreed-upon during debriefing. Low transaction costs due to build up of trust before game

8.1.2 Traffic Management System (2)

This project involved a radical overhaul of the traffic control processes in the UK. Network Rail, the U.K. infrastructure manager, found that other countries had partly automated their traffic control process and expected that a similar move would impact the reliability of their network as well. Especially the focus was to make the system more resilient to disruptions. Core of the program was to consolidate traffic control from 800, sometimes manually operated, signal boxes to 12 regional control centers, the design of new job roles and procedures for operational staff and the support through a yet to be procured traffic management system. Traffic management systems by themselves are highly singular products and the relation with the designed job roles and procedures was strong and profound.

Respondents from this project saw two distinct phases in the project. The phases were separated by a sudden push for implementation and the handing over of the project to the project department of Network Rail. Many external circumstances such as budgeting cycles and increasing pressures from the government to increase the reliability of the network created this sudden shift. Before this the project team focused mainly on studying how other countries had updated and partly automated their traffic control processes. This phase was highly exploratory looking at what to implement and how to do this.

The shift caused many of the volatility found in earlier stages to settle down for two reasons. Firstly clear institutional structures were put up, such as the decision to buy the TMS from the market, rather then to develop it themselves and to invite many parties for the tender. Secondly, in order to do so, the project team deliberately made the TMS independent from other innovation processes at that time. They intentionally created interfaces between the innovation and the incumbent system and they, to some extent, neglected potential synergies with concurrent innovations. Because of this, the complexity of managing the innovation process was mainly stemming from internal dynamics. The TMS had to be designed for a set of procedures that were not fully specified yet, while the project team was responsible for both. Contrary to the Spoorzone Delft case, were procedures were mandated by the regime, the project team's goal was to design these procedures themselves and from the ground up. Hence, whatever TMS was to be procured, the system needed to allow for these new roles and procedures. However, it remained uncertain what roles and procedures would work and be in place by 2030.

Because of this uncertainty, the TMS project team employed two gaming simulations. The first was to determine to what extent the envisioned job roles and procedures would be valuable, regardless of the TMS, and the second was to determine to what extent the systems would support these roles and procedures. Whereas the first was explanatory, hoping to confirm that the designed procedures were valuable, the latter was exploratory, similar to the SZD case.

For the first purpose, we built a paper-based gaming simulation for the Leeds area. The game (see Table 8.3) involved a realistic model of the infrastructure, the timetable and scenarios. Game players were assigned the new roles and procedures and had to solve the disruptions to best of their capabilities. The gaming simulation showed that the principle behind the new job roles and job procedures worked and provided the project team more certainty that the new roles would be valuable. This decreased the complexity as the job roles could be used as lead-functions, i.e. anchoring points, for the development of the TMS. In the subsequent development of the TMS, its value would always be evaluated in the light of adherence to the new operating procedures.

Table 8.3 LEEDS GAME.

	Usual pattern	State before game	Game design	State after game
P	Decreasing complexity in P-space due to build of momentum and interdependence from other innovations.	Remaining complexity involved designing both new procedures and TMS in conjunction	Low tech, but high realism in processes Explanatory Experimental High immersion Few scenarios	Allowed procedures to be anchored. Allowed project team to shift the focus to TMS.
S	Compartmentalization due to modularization in P-space. Low inclusion and decreasing amount of perspectives	One project team involved in both designing procedures and specifying the requirements for the tender of TMS.	Closed game design process and execution No outside observers	Little change in S-space.
I	Clear I-space due to existing project management methods. Focused on buying TMS off-the-shelve	Clear I-space allowing project team to transfer all actions from game to outside parties.		Procedures now served as method to evaluate different TMS systems.

To test to what extent the proposed TMS systems would be able to do so, the project team invested in an extensive test environment. They asked the three suppliers who were taking part in the tender to build model offices of the TMS systems to test the systems in conjunction with operators and the new roles and procedures. In three months, three teams of traffic controllers worked on scenarios in real-time in all of the three different model offices. The game resulted in 300 additional requirements. Bringing some of the operational staff that played the game to the project team helped in reducing this number to a workable 150. In some sense, the debriefing therefore went on far beyond the direct execution of the game.

To a much larger extent, this second simulation was part of the usual institutional rule setting for project management and tendering. The way additional requirements to the TMS were retrieved and handled, compared to the first simulation, was highly predetermined. The project team could simply put all requirements in the tender. The complexity of adhering to these extra requirements was then transferred to private parties. This created a commitment to action, not so much because of the game itself but more because of the institutional environment in which the game was conducted (see Table 8.4).

Table 8.4 Model Offices game.

	Usual patterns	State before game	Game design	State after game
P	Decreasing complexity in P-space due to build of momentum and interdependence from other innovations.	Procedures were anchored and served as input for testing TMS. TMS fully developed.	Highly realistic Real-time Many scenarios Exploratory	Expanded P-space by showing mismatches between procedures and TMS.
S	Increase compartmentalization due to modularization in P-space. Low inclusion and decreasing amount of perspectives	Design team of TMS (of private parties) separated from project team. Communication via strict tender procedures	Game played at suppliers. Allowed for extensive communication between project team, game players, and TMS engineers.	Players seconded from TOCs to further assist in TMS project. Separation between project team and suppliers mandated by law.
I	Clear I-space once existing project management methods were used. Focused on buying TMS off-the-shelve	Clear I-space allowing project team to transfer all resulting actions to outside parties.	No design needed for institutional structure. This was given beforehand.	No change.

8.1.3 Robust Railways (3)

Robust Railways is a program to overhaul the design and operation of the Dutch railway system. As a benchmark, Japan proved that a completely different configuration of the same technology could result in higher capacity and higher punctuality against lower costs. Hence, within certain parts of the organization of ProRail, the idea arose to reconfigure the Dutch railway system. This entailed for instance corridor separation, signal optimization, removal of railway switches, new traffic control procedures, and more reliable assets. Rather than adding-on something new, it involved the reconfiguration of existing regime elements.

The mirroring images lies in the fact that internally the innovation elements are less interdependent. Signal optimization or more reliable assets in themselves already could prove to be valuable, regardless whether the rest of the measures are implemented or not. Secondly, many of these measures are focused on elements that are also the focus of regime players. For instance, optimizing the placement of signaling alongside railway tracks has an impact on capacity and punctuality (focus of the innovation project team) as well as safety (focus of regime players). Because many of the measures are highly reversible and are less part of a web of interdependent measures, the single measures are highly influenced by external pressures. Because of these features the volatility pattern is also a mirror image of the Network Rail case. The little need to build momentum combined with the heavy external pressures later on during implementation create volatility especially at the later stages of the innovation process.

The team involved in bringing about these principles chose not to turn the innovation in one project. The overhaul of infrastructure would simply be too costly. Rather, they created interlocking mechanisms with other projects that were planned. They sought to introduce the Japanese principles at the already planned upgrades of Utrecht central station, the central hub of the Dutch network, and the Schiphol Airport-Lelystad corridor. Normally, such projects are handled using a so-called waterfall model where more strategical and longer-term design choices (such as infrastructure design and timetables) precede more tactical design choices (such as safety signaling placement) and operational design choices (such as local station layouts and traffic control procedures). However, problematic was that the innovation itself entailed coherent changes in all these layers: from infrastructure layout, to optimized signaling, to traffic control procedures.

In 2011 a project team from the traffic control department decided to conduct a gaming simulation to test the feasibility of corridor separation for traffic controllers. The renovation of Utrecht Central station, to be finished by 2015 would create more independence between two heavily used corridors that both passed this hub station. Both corridors were highly connected through railway switches. Their removal would leave fewer options for traffic controllers to divert traffic in case of disruptions. The gaming simulation, called NAU, was comparable to the LEEDS game in that it focused on testing a hypothesis. The game was a paper-based simulation using realistic timetables, infrastructure and train movements. Game players were traffic controllers, and train and personnel planners from different stakeholders as well as from different echelons. Given that the game was realistic and played in real-time, game players were highly immersed. Because the game model incorporated many different processes the design of the game was highly participative and many stakeholders were involved in the design, facilitation, and analysis of the gaming simulation session.

NAU showed that corridor separation was feasible, also for traffic controllers who now had different choices to make in case of disruptions. However, simultaneously the game also allowed for exploration of other related issues. It showed that for the measure to be effective the traffic control department needed better procedures for handling disruptions in advance as well as better communication between higher echelons of traffic control. These insights were not expected so the exploratory nature of the gaming simulation was mainly a welcome side effect. Especially the debriefing between players, facilitators and observers led to these insights. Noteworthy however is that this expansion of the scope of the project also led to discussions between the infrastructure manager and the main train operating company about the directions in which the program was heading. It therefore remained uncertain what institutions to use to coordinate actions (see Table 8.5).

Table 8.5 NAU game.

	Usual pattern	State before game	Game design	State after game
P	Rising complexity and amount of disciplines in P-space	Relative stability in P-space. Separation of innovation elements over different organizational entities.	Highly realistic game model Model contained many different processes High immersion	Decrease in complexity: game showed feasibility of innovation. Increase in complexity due to insights about additional measures
S	Boundaries of S-space increasingly fluid. From highly secluded in the beginning towards open at the end due to multiple interlocking mechanisms.	Still clear boundaries for S-space. Traffic control department solely involved in changing operating procedures for Utrecht central station.	Game model with many different processed demanded a participatory design process. Many outside facilitators and observers. Collective debriefing with stakeholders	Amount of languages remained equal. Mostly operational layers involved (from different stakeholders). Inclusion increased, more cooperative efforts instigated
I	Increasingly fuzzy I-space as multiple innovations with different I-spaces interlock.	Clear institutional structure for implementing traffic control changes	Debriefing allowed for determination of concrete actions and the coordination of these.	Unclear. Insights let to a plethora of actions, programs and projects. Controversy on direction of innovation

For the Amsterdam Airport – Lelystad corridor, the project team responsible for designing the infrastructure thought a gaming simulation would be helpful in determining which of the variants they considered scored best on managing disruptions. The variants ranged from only an extra track for overhauling at the station of Weesp, to a variant where the entire corridor would see doubling of tracks. The organization organized the gaming simulation only weeks before a letter to the government needed to be send about which infrastructure expansion was needed. Usually this would be the one that satisfied most involved actors, and with the city of Almere demanding complete doubling of tracks, this would have resulted in the most expensive solution. However, with budgetary constraints and the notion that 'Japanese' measures could help in accommodating higher traffic volumes with fewer infrastructural investments, the designers wished to test more variants.

We designed the game initially to allow for the testing of four variants. Hoping to end with one, we intended to decrease volatility in the process. However, during the design of the game many more changes in the innovation artifact become apparent. These were last minute changes in the variants themselves, additional variants and different additional innovations that could potentially be implemented simultaneously. The gaming simulation appeared to be a window-of-opportunity to test other innovations as well. For instance, with now five variants the project managers wanted to test the sensitivity of the results to the introduction of ERTMS, an innovative European-wide traffic management system, and the introduction of additional rush hour trains. For the experimental design, which begged a simple pre-test post-test, this created many problems. Firstly, the explosion of variants resulted in an infeasible factorial design: about 30 runs would be needed. The situation therefore demanded from us to make the simulation more abstract. Because game players were less immersed they were better able to evaluate all proposed measures on the spot. This resulted in a plethora of qualitative insights, which could have been very valuable for the process. However the gaming simulation was designed and conducted in a highly secluded S-space. Outsiders who had a stake in the upgrade of the railway line were not involved in the design, execution and analysis of the gaming simulation. It was therefore problematic to communicate the insights to these stakeholders (see Table 8.6 for an overview).

Table 8.6 OV-SAAL game.

	Usual pattern	State before game	Game design	State after game
P	Rising complexity and amount of disciplines in P-space	Highly volatile P-space. Many different last-minute design changes.	Intended to be explanatory. Many scenarios resulted in exploration	Little effect on decreasing complexity. Rather expanded the P-space.
S	Boundaries of S-space increasingly fluid. From secluded in the beginning towards open at the end due to interlocking mechanisms.	Expanded S-space. Involvement of many different disciplines and languages: infrastructure managers, municipalities, train companies.	Expanded S-space not involved in design and execution of game because infeasible to determine who should be involved. Low inclusion.	No change in S-space. Therefore expansion in P-space was not met with actors in S-space willing to act upon the results.
I	Increasingly fuzzy I-space as innovations with different I-spaces interlock.	No clear institutional structure. Many, but weak, ties between actors	I-space not incorporated	No change in I-space.

8.1.4 Synthesis

Gaming simulation is able to cause convergence and divergence. In some instances the gaming simulation opened up the P-space by showing additional complexity or by bridging different disciplines. Furthermore gaming simulation was able to allow for outsiders to enter the S-space, creating diversity in languages and perspectives. In other instances, gaming simulation did the opposite. It allowed project managers to contract the P-space, decreased complexity and the amount of disciplines. Furthermore, it could show what actors were needed for successful implementation of the project and hence allow the S-space to be closed for others.

Looking at the dynamics in P, S and I spaces beforehand showed us that the relation between game design and effects on volatility are highly history dependent. This is because for explanatory games, which should test hypothesis rather than generate them, game designers need a clear and stable P and S space. However, in instances where convergence is needed to counter the high volatility the P and S space are inherently unstable. The OV-SAAL game perfectly shows how gaming simulation was unable to allow spaces to converge. In the instances were convergence was realized (LEEDS and NAU), this was ensured by having the game focused on specific subset of an innovation. This made it easier to determine the game model and the actors needed to create commitment to action.

For exploratory and divergent purposes of gaming simulation to be valuable it needs an expanded S-space to co-align with the expansion of the P-space. At the Delft case this was ensured by the installation of the commissioning team prior to the execution of the game. Given that the game was designed, executed and debriefed by this commissioning team the already existing trust between stakeholders increased the commitment to action. Therefore the many qualitative insights were countered with effective measures. For the MODEL OFFICES game, the expansion of the S-space took place during the gaming simulation and the already designed I-space for handling the results leaded to commitment to action.

8.2 Gaming and mechanisms

From the five games we used for the analysis in the previous part, we now add another four that we designed and employed in the railway sector and focus more on how mechanisms were enabled or blocked during the design and execution process of the game. We leave one game out of the scope. The MODELOFFICE game from the Network Rail case was a game where we were not involved in designing it and therefore prohibits us from directly determining what mechanisms were involved. Hence, the analysis will revolve around 8 instances of the employment of gaming.

Mechanisms translated into game dynamics

For the analysis we have to more fully operationalize the four mechanisms and translate it into an analytical framework more specified to gaming simulation. In Table 8.7 we provide an overview of what the mechanisms might mean during game design, game employment and the debriefing:

Table 8.7 Relating mechanisms to gaming simulation

Mechanism	Games' impact
Ambiguity	Do trade-offs appear to a broader set of stakeholders than the immediate involved project members? Does the functionality of the innovation become clear? Do impacts of the innovation on other processes become clear?
Disentanglement	Does the structure of the innovation become more clear? Does the innovation become operationalized? Do dominant parameters of the current system become clear in relation to the innovation?
Interlocking	Do projects become intertwined over the course of the game Do results indicate further action beyond the immediate remit of the involved stakeholders?
Shielding	Openness of the game, are new stakeholders introduced due to game design decisions? Do these stakeholders bring with them influences to the innovation processes that might be described as regime pressures (e.g. long-held norms and rules) Does the game leave intact the assumptions that designers used before the game?

8.2.1 Games

All simulation experiments involved solely the P-space as the simulant (see Table 8.8). Innovation actors either explored systems that they would design later on, or tested innovations already designed. There were no gaming simulations to explicitly, in game, explore the S and I-space. Those games would for instance involve the simulation of design processes, rather than operational processes, and through this learn how to improve such collective processes. This does however not mean that the dynamics in S and I spaces played no role before and during the game nor does it mean that these spaces were not affected. As we will show later on, they had a serious impact on the feasibility to open up or close down P-spaces as well as were significantly changed by the experiments in the P-space. Secondly, the table shows that it is not self evident that games designed to do one thing will result in doing exactly this thing. At Bijlmer, NAU, and OV-SAAL we saw that in the end the games caused divergence while the purpose of the whole exercise was convergence. Table 8.8 provides an overview.

Table 8.8 Overview of gaming simulations and their intended and realized impact on P-space

Game	Goal	Model	Intention	Impact
BIJL-MER	Testing predesigned traffic control procedures for dealing with future high traffic volumes around Amsterdam	Paper-based model of infrastructure, highly detailed computerized interfaces, realistic timetables, traffic controllers	Closing P-space	Opening P-space
ETMET	Testing predesigned traffic control procedure for dealing with disruptions on the Dutch central corridor and metro-like timetables.	Paper-based model of infrastructure, low tech interfaces, realistic timetables, traffic controllers from different organizational entities	Closing P-space	Closing P-space
NAU	Testing predesigned traffic control task separation to unbundle areas of control around central hub of the Dutch network	Paper-based model of infrastructure, low tech interfaces, realistic timetables, traffic controllers from different organizational entities	Closing P-space	Closing / Opening P-space
WINTER	Replay of traffic control processes when network gridlocked due to snow	No technical model, communication between operators disciplined by real timeline of events, traffic controllers as players	Opening P-space	Opening P-space
LEEDS (UK)	Testing predesigned traffic control roles and procedures needed for consolidating traffic control from 800 local control centers to 13 regional centers.	Paper-based model of infrastructure, low tech interfaces, realistic timetables, traffic controllers from different organizational entities	Closing P-space	Closing P-space
TMS (UK)	Finding requirements for the design of a traffic management system.	High tech prototypes of traffic management system, realistic timetables, traffic controllers	Opening P-space	Opening P-space
IPAT	Finding additional requirements for the design of a railway tunnel	High tech prototype of tunnel hardware and software, representatives and operators from operational echelons as players	Opening P-space	Opening P-space
OV-SAAL	Testing four predesigned infrastructure expansions on their robustness against medium-sized disruptions.	Paper-based model of infrastructure, low tech interfaces, realistic timetables, traffic controllers from different organizational entities	Closing P-space	Opening P-space

8.2.2 Ambiguity

Of the four mechanisms used in the analysis of the role of gaming simulation, ambiguity was the most problematic. Many of the innovations that were tested in the gaming simulations were not of a systemic nature and hence potentially did not benefit from any ambiguity. Also, some were unambiguous to begin with. For the analysis we therefore look solely at

NAU, 1st PHASE and OV-SAAL. In these instances, the innovations started out as more abstract notions that in the arena of involved stakeholders could be ambiguously interpreted. We see that during game design, run and debriefing this ambiguity is alleviated (see Table 8.9)

Table 8.9 Gaming simulation and ambiguity

	Game Design	Gameplay	Debriefing
BIJLMER	No effect. Innovation itself was already unambiguous	No effect	No effect
ETMET	No effect. Innovation itself was already unambiguous	No effect	No effect
NAU	Decreased. During design process functionality of innovation appeared to be perceived differently between two stakeholders	Decreased, immediate impact of innovation became apparent during game run and observers from stakeholders were there	No effect
POP	No effect	No effect	No effect
1st PHASE	No effect	Dilemmas appeared to broader set of stakeholders as game players went on to operationalize the innovation during game run (see disentangling)	No effect
WINTER	No effect	No effect	No effect
OV-SAAL	No effect	No effect	Addition of stakeholders from different disciplines broad to light the inherent choices that the game was looking for
IPAT	No effect	No effect	No effect
CONOPS	No effect	No effect	No effect
		Gameplay	Debriefing

For NAU there appeared to be a difference in perception of the intended functions of the innovation between two key stakeholders. For the train control department of Dutch railways the New Actionplan Utrecht was about diminishing workload for their operational echelons and for ProRail it was about increasing the robustness of the central node of the network. So before the game was initiated the cooperation between the two stakeholders was based on the (partially false) understanding that for each party certain benefits were to be realized and the game caused more clarity in what each party was to gain. Although the innovation drivers assumed decreased workload was a welcome side effect, their design choices were not focused on achieving this. Also, whereas in the beginning the actionplan was highly abstract and hence did not portray all immediate dilemmas, the running of the game showed that under the new regime of removing switches and appointing traffic controllers to fixed areas, large disruptions were to be expected when certain key parts of the infrastructure would malfunction. These dilemmas were inherent to the innovation but remained unclear because of the ambiguous nature of the innovation before the game. Given that the game run was open and many stakeholders send out observers to see the game run, the ambiguity further decreased.

1st Phase was similar in that the traffic control concept was still in its developmental stage and hence of an abstract nature. However, different from NAU this game had a more closed design process. So the dilemmas that popped up during the design process, if any, remained inside a small subset of relevant stakeholders. Ambiguity did decrease during the game run but to a lesser extent than was the case for NAU. This had to with the fact that game players had to further operationalize the innovation on the go but fewer external stakeholders were there to observe the dilemmas this brought about.

OV-SAAL is unique in that the ambiguity was solely alleviated in the debriefing stage. The game design process was so that stakeholders could enter the process and influence what innovations were to be tested (see interlocking later on) but information on the core innovation did not leave the design process. Comparing this process to the design process of NAU, it was much more unidirectional in the way a broader set of stakeholders interacted with the game designers. Also, the results of the game were to be shared with new stakeholders during the debriefing. Hence it was purposefully setup to keep the game design process and the game run more closed from external stakeholders and only present the results to them in the debriefing. In the debriefing we presented the results to stakeholders like municipalities who now saw the direct consequences of the innovation that was tested. Given that one consequence was less use of additional infrastructure and cities like Almere were, for other reasons, in favor of the doubling of tracks within their city limits, this significantly decreased ambiguity.

8.2.3 Disentanglement

Innovations usually start as abstract notions or vague theories on how a current system can be improved. As we have noted in the previous chapter for an innovation to be implementable it needs to be disentangled. This disentanglement is an interactive process between finding the dominant parameters in the current system and detailing the innovation so that it becomes operationalized. An overview on how different games ensured or blocked disentanglement is given in Table 8.10.

Table 8.10 Gaming simulation and disentanglement

	Game Design	Gameplay	Debriefing
BIJLMER	Partially, failed beforehand	In-game further operationalization of innovation by game players, but lack of immersion led to expansion of design space	Discussion moved away from innovation but focused on other projects that could help traffic control
ETMET	Game design process forced stakeholders to operationalize their innovation. Partially succeeded	In-game further operationalization of innovation by game players. Worked because they could rely to some extent on old ways of working	No effect
NAU	Full disentanglement	No effect	No effect
POP	No effect	Innovation designed during game (design game)	Further discussion on workability of solution
1st PHASE	Partially, but innovation not fully operationalized before game	Game became partly an experiment and partly a design exercise	Unconstructive debriefing, mostly focusing on the method.
WINTER	No innovation gamed (diagnosis game) so no effect on operationalization of innovation	Game involved a stepwise replay of one day, high immersion and little discussion on what the core problem was.	Collective discussion showed the dominant parameters of the system and avenues for further improvement
OV-SAAL	Full disentanglement before game, but still many changes to innovation during game design	No effect	Addition of new stakeholders in debriefing created dynamics in already operationalized design.
IPAT	Technical part of tunnel fully disentangled beforehand. Gaming demanded procedures to be more fully specified than was the case.	No effect	Rich discussion on further design of the conjunction of tunnel and procedures, led to more operationalization
CONOPS	Already disentangled before game design started	No effect	No effect

We see that only two games, IPAT and WINTER, led to disentangling the innovation in the debriefing. Especially the latter game, which did not involve any innovation but was merely intended to provide a diagnosis of the current system, is unique in that sense. WINTER provided stakeholders an opportunity to collectively observe holistically a cumbersome process (traffic control during wintery weather and many disruptions) and find the dominant parameters. Only during debriefing was this possible as the game itself was played semi real-time and players and observers were fully immersed in the game. Although no innovation was to be operationalized, the finding of dominant parameters (that the problems were of structural nature and not due to traffic controllers individual behavior) significantly helped in fine-tuning any remedy that was being discussed parallel to this game.

IPAT saw a similar debriefing, but now with a concrete innovation to be discussed. While the technical part of it was fully developed and prototyped in the game model, the procedures to operate the tunnel were not fully operationalized yet. Part of the reasons the game was played was to front-load these design process to a stage where technical changes were still possible. During the debriefing a more fully operationalization of the entire artifact (technical part and the procedures) was possible. The proximity of otherwise dispersed stakeholders and the collective observation of the game run allowed for a quick discussion of additional measures that were needed, measures which were inside the scope of the involved stakeholders so as to no invoke any interlocking processes.

In most of the other games we see that disentanglement is mostly done in the game design process or is already done beforehand. This has to do with the nature of gaming, since it demands a somewhat workable innovation to be tested in a simulation environment. At CONOPS the innovation was fully detailed by the gaming client so the game did not have any effect on disentanglement. The other games show an interesting phenomenon. During the design of the game, innovation stakeholders are usually pressured to make their innovation more concrete. We then see two choices: fully operationalizing it, as is the case for NAU and OV-SAAL, or partially operationalizing it and hoping that game players are able to 'fill in the blanks'. For BIJLMER, 1st PHASE and ETMET we then see that this reliance on game players to further operationalize it during the game run can be problematic. BIJLMER and 1st PHASE were games where the ambiguous nature of the tasks (playing the game and operationalizing the innovation at the same time) for game players led to immersion issues. For ETMET this proved not to be a problem. We assume this has to do with the fact that that part of the innovation that was left to the decision of the game player stayed somewhat close to their daily routines and behavior.

Disentangling an innovation during the game run is in itself not inherently problematic but demands careful game design. The goal of POP was to do just that, enable game players to design procedures to make use of idle capacity around railway stations during the night to park trains. There was an abstract idea about a certain process to achieve this goal, but it was left to game player to fully design the process in detail. Hence, the game was specified as a design game and this was clearly communicated to the game players. Although some players had issues with playing the game and designing the process at the same time, it places demands on their cognitive abilities, the goals they had to accomplish were clear. In the end this game resulted in a more detailed process for routing trains during the night.

With disentangling an innovation comes an increased focus on details. The scope of the innovation under study tends to decrease and the specific focus tends to be the result of a path dependent process. Initial design decisions may heavily steer the development of the innovation towards a specific end result. In Chapter 7 we showed how the analysis of the Dutch railway system and the Japanese railway system between 2005 and 2009 heavily impacted each other. These same might be true for disentangling an innovation in-game. There seems to be a natural tendency to both play the game and further detail the innovation. The same path dependent process can impact this detailing as we saw in innovation processes that games should support. At POP the question is if the end-result is most optimal or merely a logical result from game players decisions at the beginning of the session. This is a worrisome feature of games. Being inherently a group activity it might be endangered by groupthink: too soon a focus on a single solution. Hence, such problems must be alleviated in the debriefing. Were early decisions in the game impacting subsequent decisions? And what would the end-result be had these early decisions been different? In Chapter 9 we devote some attention to the use a qualitative robustness analysis on the game results using the debriefing.

8.2.4 Interlocking

Innovation processes rarely occur in isolated environments. They exist in an eco-system of other projects. The innovation that is tested in a game might connect with other projects or other projects might attach themselves to the innovation under study. In general we see that games can function as platforms for innovations to connect, as shown in Table 8.11. This can be helpful to give an innovation more momentum but it can also be detrimental as an innovation looses autonomy.

OV-SAAL is an interesting game since the game appeared to be a window-of-opportunity to test other innovations as well. At the start of the game design process the initial goal was to test four different infrastructural variants of the Amsterdam-Almere corridor. However, during the process other stakeholders asked us to incorporate additional measures. In conjunction with the infrastructural variants, the introduction of ERTMS and new timetables with added rush hour services were to be tested. This resulted in a dilemma since we had to somehow still be able to discern the impact of the infrastructure variants alone (which was the demand from the initial client) and in combination with the other measures. During the game run it appeared that the design choice we made to do so, opting out of using realtime play to simulate many variants in shorter periods of time, resulted in loss of immersion for game players. Given that they were less immersed they pointed to additional measures needed, such as a track for shunting trains midway between Amsterdam and Almere.

The relationship between immersion levels and interlocking processes is also profound in other games. At BIJLMER we saw that due to loss of immersion game players were more able to pinpoint other measures needed to accommodate the introduction of the innovation. Not fully behaving as they would in real life, they were better able to co-design the innovation. This was not the initial goal of the game and proved to be difficult to deal with. The resulted interlocking processes, with future projects such as platform extensions and overhaul tracks, was problematic given that the needed stakeholders for those projects were not involved.

Table 8.11 Gaming simulation and interlocking

	Game Design	Gameplay	Debriefing
BIJLMER	No effect	No effect	Increased due to discussion of other needed measures to accommodate innovation
ETMET	No effect	No effect	No effect
NAU	No effect	No effect	Increased due to additional measures related to innovation being discussed
POP	No effect	No effect, design boundaries were communicated to players beforehand	Increased due to additional measures needed to accommodate the innovation that was designed in-game
1st PHASE	No effect	No effect	No effect
WINTER	No effect	No effect	No effect
OV-SAAL	Game functioned as window of opportunity to test additional innovations	Game players pointed to additional needed measures to accommodate innovation	New stakeholders entered during debriefing and added other innovations to be tested or discussed
IPAT	No effect Project was properly demarcated beforehand	No effect	No effect
CONOPS	No effect	No effect	No effect

NAU was similar in that during the debriefing the discussion was mainly about additional measures to be taken to make the innovation under study more effective. During the game run observers and game players saw that the innovation caused little flexibility for the traffic controllers in Utrecht. Even though disruptions were happening in Utrecht, these had to be managed on a larger geographical scale. This led to a discussion on the workings of the national traffic control center, a newly introduced innovation at that time, and what their role had to be when the innovation was implemented. Hence, the NAU project interlocked with processes on redefining the national traffic control center. This was less problematic since most of the involved stakeholders were already there during the game run. Still however, the way this coupling of projects had to be realized was unclear, even after the debriefing.

8.2.5 Shielding

For innovations to be explored the involved stakeholders need to be able to shield themselves from the immediate pressures from the regime system. This shielding is needed because systemic innovations might be beneficial once fully implemented but are met with hostility by the regime. However, our analysis also showed that shields tend to persist resulting in the use of invalid assumptions and creating technical interlocking processes without the needed social interlocking for coordinating these interlocked projects. The extent to which games allow for shielding is shown in Table 8.12.

Table 8.12 Gaming simulation and shielding

	Game Design	Gameplay	Debriefing
BIJLMER	No effect		
ETMET	Decreased due to needed involvement of many stakeholders to design game	Decreased due to involvement of operational echelons as game players	No effect
NAU	Decreased due to needed involvement of many stakeholders to design game	Decreased due to involvement of game player and observers from different departments	No effect
POP	No effect	Decreased due to involvement of game players	No effect
1st PHASE	No effect	Decreased due to involvement of game players	No effect
WINTER	No effect	No effect	No effect
OV-SAAL	No effect	Decreased due to involvement of game players	Decreased due to purposeful introduction of external stakeholders
IPAT	No effect. External stakeholders had been introduced to the process beforehand	No effect	No effect
CONOPS	No effect.	No effect	No effect

Games inherently alleviate these shields, as they need the involvement of operational echelons for playing the game to be realistic. In general we see that the introduction of players that are used to behaving in the context of the current system (they have experience, mental models and heuristics that allow them to deal with the complexity they are faced with in their daily work) can be problematic when testing an innovation that is systemically different. Cognitively they alleviate the shields since the game players can show directly and indirectly how assumptions held by designers of the innovation are valid or invalid.

At NAU and ETMET shielding was diminished because the game design processes demanded the involvement of a plethora of new stakeholders. Although traffic control had worked out a concept for dealing with increased train traffic under the PHS timetable, the design space bordered closely or even overlapped with the design space usually assigned to other departments. For disentangling the innovation to be made workable in the game, they had to incorporate previously external stakeholders.

8.3 Game design parameters and patterns

Our assertion from Chapter 6 was that certain macro-level patterns are unique to systemic innovation processes and may sometimes be considered problematic. The pattern we found and which we related to the underlying mechanisms in the subsequent chapter is that dynamics in product, social and institutional space tend to increase rather than settling down. As we proposed, gaming simulation's main value could lie in front-loading some of this volatility to earlier stages of the process. In this paragraph we try to find how different design parameters of a game, e.g. model representation, type of game players or the use of real time, impact its ability to control dynamics in these spaces. We again use the games we designed for the railway sector.

8.3.1 Expanding the P-space

IPAT, WINTER, and TMS were games that were designed to open up the P-space. Of these three WINTER was the only game solely meant for diagnosis purposes. This game was designed to allow operational personnel of the Dutch railway sector, such as train and traffic controllers, to replay a day where the whole network collapsed due to wintery weather. Before the simulation was conducted, the designers of procedures for handling disruptions knew that the low robustness and resilience of the network was not caused by individual behavior nor solely by technical failures. Rather they wanted to explore where in the cooperative structures between different operators the reasons lie for the system's inability to cope with train and track failure down due to icing. For representatives of higher echelons this presented a chance to study processes holistically that are otherwise separated in space and time.

The other two games were designed to specify requirements for technical artifacts in conjunction with operator roles and procedures. IPAT intended to deliver a set of issues revolving around the mismatch between the software and hardware of a tunnel and the procedures designed to operate it. TMS was a test to study what changes had to be made to the design of an intelligent traffic management system to allow it to support future operational roles and procedures in the British railway sector.

For all three of them hold that before the employment of the game no hypotheses were present. Because of this the game model had to be large enough to allow for exploration. This entailed both a large number of processes from the referent system becoming part of the game model and the running of many scenarios. By doing so, we decreased the chance of overlooking certain interesting aspects of the real life system. Furthermore we found that the ability of a game to allow for exploration relies heavily on the dialogue between operators (of the P-space) playing the game and designers (from the S-space) observing the game. Real-time play and immersion seem undesirable. Both parameters of game design inhibit game players to be in a reflexive mode, a mode that is crucial for creating the needed dialogue. In addition, the IPAT ensured this dialogue by having representatives from the S-space with sufficient operational knowledge play the game thereby effectively ensuring a dialogue between the P and S-space inside one person. This feature helped greatly in translating the outcomes to concrete actions afterwards.

Of crucial importance for effectively designing exploratory games is that the many insights that result from it are acted upon after the game. Therefore, next to expanding the P-space, the S and I-space need to expand as well. For the IPAT game, the S-space was already expanded due to the instigation of a special commissioning team with representatives of the tunnel project and the current organization. This allowed for the design of the game to incorporate all relevant processes, as well as for the design of interesting scenarios. This open design process made sure that the exploration during gameplay would touch upon all factors deemed relevant by all stakeholders. Additionally, the institutional structure was already in place, making it for most of the requirements easy to determine who was responsible for what. The build up of trust well before the employment of the game also caused game participants and observers to design new institutional arrangements in the debriefing. This was especially valuable for those requirements for which the specific coordination mechanisms were yet unclear. The TMS game was similar in that the way requirements were acted upon was strictly organized through market mechanisms. The simulation results could simply be put in the upcoming tender, as requirements, for which three suppliers of the TMS were still in the race.

8.3.2 Closing the P-space

When a simulation outcome rejects or accepts a hypothesis, in this case about the effect of an innovation, the P-space contracts. Knowledge availability increases and the innovation can now be implemented by dividing the innovation in modular work packages. This decreases the complexity and makes it less multidisciplinary. The ability of gaming simulation to ensure this effect relies on the perceived validity and reliability of the outcomes of the game by all involved stakeholders.

To provide with more certainty that an innovation has a certain effect on for instance punctuality, robustness or resilience, the gaming simulation needs to be designed according to strict experimental design principles. Stakeholders with which we designed and executed successful games, such as ETMET, NAU and LEEDS, all deemed these valid because of a range of similar features of these simulations. Firstly, they often involved a pretest-posttest experimental design, creating higher internal validity. Secondly, a clear conceptual model on the links between innovation and a predetermined dependent variable was present. Therefore we were able to operationalize the parameters we are interested in as well as structure the debriefing and analysis of the game in such a way that hypotheses could be accepted or rejected. Thirdly, we modeled processes from the referent system in high detail. Stakeholders often acknowledge the sensitivity of overall system behavior to small changes and only processes with high detail can replicate this behavior. Realistic movements of trains according to real life timetables, precise procedures for communicating between operators are a few design parameters by which we ensured external validity. This high detail creates the need for many different processes to be modeled in the game because system boundaries expand and ecological validity of the game needs to be maintained. Furthermore we often use real-time play to allow for this high granularity in processes to become valuable. Fourthly, immersion of game players is of vital importance. Whereas for exploratory gaming simulations we need a dialogue between operators and designers, in these explanatory gaming simulations we need operators to act precisely as they would in real life. Real-time play, or at least time pressure, and high detailed processes seem to contribute to immersion. On the other hand we have found that the relation between level of detail of representation seems hyperbolic. With our very low-tech representations (infrastructures printed on whiteboards, sponges as trains) we have seem to create higher levels of immersion than games using more high tech and detailed interfaces (BIJLMER).

Gaming simulation as pure experiments have many shortcomings. Often we can only run a few simulations in one a day, threatening the reliability of our results. Usual ways of overcoming these such as repeated runs, sensitivity analysis and elaborate factorial designs (Balci 1998; Sargent 2005) are therefore infeasible. In previous work we provided a framework for the debriefing of such explanatory gaming simulations to alleviate some of these validity threats (Van den Hoogen et al. 2014). Here the debriefing should allow for the assessment of the reliability and sensitivity of the outcomes.

To effectively contract the P-space the relevant S-space should at least observe the gaming simulation and participate in the debriefing. This way, stakeholders can observe why an innovation brought about changes in the dependent variable. This increases the confidence the S-space has in rightfully contracting the P-space. Next to that, we found that the S and I-space deserve little attention. This is because the game design process already incorporates the relevant S space and often the innovation to be tested is already on the way to implementation. For instance in both NAU and ETMET the game model incorporated many different operational processes creating the need to incorporate designers of these processes to be involved in the game design process as well. Next to that, the contraction of the P-space leads to the effect that rarely actors outside of the current S-space need to act upon the results. In other words, if a game is about traffic control the results will not impact the design of safety signaling.

The only exception of this is OV-SAAL. This was because the game was about testing four variants on one dependent variable, whereas the final choice for the variants incorporated many other variables. In these instances a much broader S-space must be incorporated in the design process of the game. However, given that these designers usually stem from completely different disciplines and the specific P-space for the game only revolves around one aspect system, this is hard to ensure

8.3.3 Mixed results

BIJLMER, NAU and OV-SAAL were games that had a different impact on volatility then originally intended. The causes for this we propose are faulty design choices, the context in which the game was employed as well as a natural tendency of collaborative simulation efforts towards exploration rather than explanation. In a sense, this is not inherently a bad thing. Additional insights can prove to be highly valuable. For instance, the NAU game showed the directly involved innovation managers that the traffic control procedure they had designed worked well. On the other hand, the game also led to insights about additional measures that had to be taken for the innovation to fully work. The fact that this game had many interrelated processes in the game model led beforehand to the involvement of many actors in the design process. Therefore the S-space was already expanded before the actual employment of the game and many observers from different organizational entities were present during the run. This created more possibilities to act upon the expanded P-space. However, the I-space was still uncertain. The actual coordination mechanisms as well as the directions in which the innovation was going became a hot topic of debate, during the debriefing and long after the game was finished.

The BIJLMER game saw a similar dynamic. However, here the P-space was expanded to such an extent that infrastructure design and station layouts became part of the solution space. This unexpected expansion of the P-space was not met with a coherent expansion of the S-space as nobody expected these kind of aspect systems to become part of the P-space. This led to many insights not being capitalized through the concerted actions of many actors. This in contrast to games that were intentionally designed to expand the P-space where the involvement of the S-space in the design and execution of the game (IPAT) or the careful design of the I-space beforehand (TMS) led to a coherent change in multiple spaces.

8.3.4 Uncanny valley

The BIJLMER game was one of the first games we designed for the railway sector. The game served to test a traffic control concept that was deemed necessary if in the future traffic volumes were to drastically increase. In the game model we tried to achieve high detail in processes and interfaces, because we intended to make the results internally and externally valid. However it seemed that players had problems with the accuracy of the interfaces. This created low levels of immersion and posed threats to the validity of the simulation results. Comparing to later simulations we have designed, it appeared that an uncanny valley effect can arise when designers strive for too much similarity between the model and the referent system. Slight difference between the game and reality then significantly impact the way game players experience it. In other similar games, lower tech representations often worked much better.

8.3.5 Context dependence

The OV-SAAL game provides a perfect example where the context in which the game was conducted led the impact of the game to be completely different from what was originally intended. In a highly volatile and political context, we were asked by the traffic and logistics department to design a simulation of the Amsterdam Airport – Lelystad corridor. This corridor was to be upgraded and the department had four variants they wished to analyze on their robustness. Robustness is the extent to which the infrastructure gives the traffic controllers enough possibilities to cope with small and medium-sized disruptions. We designed the game as an explanatory gaming simulation, hoping that the outcomes would be a convergence on one of the four variants.

However the highly volatile environment in which we designed and executed the game led to a few interesting results. Dynamics in the P-space before and during our design process resulted in many changes in the variants, even during the game. Additionally the gaming simulation provided a window- of-opportunity to test the variants in conjunction with other changes such as newly designed timetables and a capacity-increasing safety system. These dynamics led to an explosion of our factorial design. Our initial desire to simulate in high detail, and probably in real-time, became impossible because we needed 20 runs. Operators are rarely available for longer than a day and the results were to be delivered quickly. Because the game now became more abstract and game players were less immersed the analysis of the many variants became highly qualitative. It also expanded the P-space because of a rich dialogue between facilitators who observed the simulation and operators who played the simulation. However when trying, in the debriefing, to converge on one of the variants that according to game players was most optimal, we found that the designers who were only present during the debriefing found this variant no longer relevant. So in highly volatile contexts gaming simulation sees two threats: firstly, volatility impacts the ability to design simulations in such a way that they converge on one solution. Secondly, the delay between the question (as input of the game) and answer (as output) means that these do not align anymore when P-spaces are highly dynamic.

8.3.6 Eigendynamics

Gaming simulations bare in them an internal tendency to create divergent and exploratory processes. The fact that we bring together operators (part of the P-space) and designers (from the S-space) creates certain expectations. For innovation managers this is one of the few times they actually communicate with operators. And for operators it is one of the few moments they are incorporated in the design process. Especially since they themselves are part of the to be designed product, they see these gaming simulations as an opportunity to influence the design process. These expectations from both sides create an internal force that pushes towards dialogue between these separated worlds. However to ensure convergence we need the simulation run to be externally valid. This demands from game players to act as they would do in real life. A constant dialogue between operators and designers is certainly not part of this real life. Other validity threatening aspects of gaming simulation play a role in its inherent problems of ensuring convergence. Humans participate in gaming simulations and they not always behave as experimentalists would like them to behave. They get distracted and sometimes do not follow exactly the rule-set designed beforehand. This is problematic as it hampers reliably coming to one valid conclusion.

8.3.7 Design parameters

From the many games we have designed for two different purposes we distill a few design parameters that made them effective. For the game designers these parameters enable a careful manipulation of the effect of a gaming simulation. In Table 8.13 we provide an overview of these parameters.

Table 8.13 Game design parameters

Parameter	Exploratory Games	Explanatory Games
Experimental design	Single test (per scenario)	Pretest Posttest
Amount of processes	As much as needed to find interesting phenomena	As much as needed to ensure ecological validity
Process detail	Low	High
Measurement	Flexible methods and sources	Predesigned measurement instruments
Immersion	Low	High
Game players	Players with operational knowledge	Real operators
Real-time	No	If needed
Scenarios	Many	Few
Flexibility	Yes, allows for searching and finding of interesting phenomena	No, endangers internal validity
Dependent variable	No, might emerge from game	Predetermined, fully operationalized
Debriefing	Unstructured, focused on insight	Structured, focused on validity

In two instances the game design process itself needs careful consideration by the game designer. When designing explanatory gaming simulations in highly volatile times the design process can alleviate many of the problems found for these games. An open and flexible process allows the game design to move with the volatility in the P-space. Last minute innovation changes can quickly be incorporated in the game. This ensures that the effect of the lag between game design and outcomes is reduced.

For exploratory gaming simulations, the design process is much more important. The simulation should explore a vast problem and solution space and beforehand its boundaries are unknown. The incorporation of many innovation actors increases the chance that the game touches upon a wide array of interesting phenomena. For instance at the IPAT game everybody was able to contribute to the design of scenarios, making sure that after the gaming simulation all relevant phenomena were uncovered. Additionally, the game design process is the moment that changes in the S and I-spaces can be realized. These are needed to allow the insights from the game to have any impact outside of the game. An open process results in a joint fact-finding session, where different actors can discuss the model and the assumptions. These features of an open design process make sure that all actors share the results.

8.4 Conclusion

This chapter identifies multiple roles of gaming simulation in innovation processes. Over the course of an innovation, games have shown to have a strong interaction with process volatility, and therefore provide more functions than the traditional simulation function of generating knowledge. This is in line with Duke and Geurts' (2004) work on policy interventions with gaming, but now also shown for innovations.

Gaming simulation design decisions can ensure the manipulation of volatility by allowing for the opening and closing of product spaces. This chapter has provided a set of design parameters by which game designers can create either of the two effects. Game model detail, immersion levels of players, amount of scenarios and experimental designs are some of parameters designers can manipulate. One design parameter that influences this effect of gaming simulation appears to be the rigor by which it is employed as an experiment. Few scenarios, high immersion, clear dependent variables and the use of realistic models increased the extent to which the game could bring convergence in designs and perspectives. On the other hand, exploratory effects were realized using a multitude of scenarios, lower immersion and more communication between players and observers during gameplay. The insights that exploratory gaming simulation can deliver are only capitalized on when observers are used and an extensive debriefing takes place.

A more detailed look into the working of gaming simulation involved the use of dynamic mechanisms that play a role in driving systemic innovation processes. Games tend to alleviate shields, ambiguity and increase interlocking. Also, games force designers to more fully disentangle their innovation since workable models are needed. If left partially open, gameplay will further disentangle an innovation. This warrants however close attention since the operationalization of an innovation now is partially out-sourced and occurs in-game. As long as the mechanisms can be seen as constructive, these influences of the use of gaming can be seen as valuable. However, as Chapter 7 showed, the mechanisms bear in them a constructive and disruptive side. Some of the games indeed allowed for too much interlocking and alleviated an ambiguity that could have been constructive. For designers of games these considerations are important.

Furthermore this chapter pointed to further pitfalls gaming simulation designers can encounter when designing games for innovation processes. Because gaming simulation needs the involvement of many disciplines and encompasses the bringing together of operators and designers they have a tendency to open up the P- space. This has three implications for game design. Firstly, for game designers wishing to design an explanatory game this means that more energy is needed to encounter this tendency. Secondly, to capitalize on expanded P-spaces, game designers should take into account the specific constellation of the S and I-space. Otherwise the many insights that games deliver are not translated into coordinated actions. Thirdly, the context in which gaming simulation is employed impacts to what extent a game designer can direct a game towards closing spaces. Especially in already volatile situations, where dynamics in P, S and I spaces are profound, it is a cumbersome task to design explanatory gaming simulations. Volatility begs for explanatory gaming simulation, due to its ability to close spaces, but volatility itself forces games to become ever more exploratory. This points to the inherent weakness of gaming simulation as a pure analytical-scientific instrument. This point was already made in Chapter 3 but the findings here corroborate this finding.

The active substance of gaming simulation, when employed in innovation processes, is the manipulation of volatility. In several cases the method has shown its ability to either front-load volatility or to decrease volatility. Especially the latter effect is not uncontroversial. We have seen in many instances gaming simulation decreased complexity in the P-space for one specific part of the innovation but opened up a can of qualitative insights about other elements. This effect seems almost inherent to the use of gaming simulation, but if not expected by those who employ the method, the expansion of the P space will not be capitalized on. Expansions of P-spaces, as we have seen, are only valuable when S and I spaces are either carefully designed beforehand or are similarly impacted by the game design process and the game itself. This provides additional challenges to the design of so-called exploratory gaming simulations as for to be valuable it is not solely about the design of the experiment but also about the design of the game design process, the session itself and the debriefing.

In contrast, our analysis in Chapter 3 dealt with the inherent value and shortcomings of gaming methodology, irrespective of its use in a specific context but perceived through two models on innovation and gaming (see Thomke, 2001 and Klabbers, 2009). Validity issues are prominent, only certain search strategies are possible and transition planning and joint fact-finding can be supported using gaming. What this picture shows is that games can be flawed but potentially be improved by better designing, facilitating or debriefing the gaming session. However, the subsequent analyses shows that games value is much more complex since it involves manipulating mechanisms and patterns that bear in them constructive and disruptive influences at the same time. Sometimes less valid games might be more suitable since they can leave the ambiguity of an innovation intact, sometimes joint fact-finding might be detrimental as it will increase the interlocking between projects too much or alleviate the niche-like qualities of the initial innovation actor arena. Furthermore, a more careful transition planning beforehand might show all the inherent dilemmas of systemic innovation at a time when dealing with these dilemmas is not yet opportune.

Games have inherent shortcomings and shortcomings due to contextual influences. For instance, given the tendency of gaming simulation to cause exploration, it becomes especially important for games that intend to cause convergence to find ways to counteract this tendency. We have seen that such gaming simulations led to divergence instead through a less structured debriefing. In light of this, a structured approach to the design of the debriefing might be valuable. In addition, validity issues, specific search strategies, a more careful transition planning as well as a controlled manipulation of relevant mechanisms such as ambiguity and interlocking can be done in this phase of the gaming simulation session. Whereas the simulation in itself is usually to some extent a laissez-faire affaire, especially when played in real-time, the debriefing is the moment where all these issues can be intentionally tackled. Hence the next chapter deals with designing such a debriefing.

9 A Normative Framework for Debriefing: context, substance and method

Although the role of gaming is much different from being solely an experiment, the results of a gaming simulation should still be grounded in some perceived reality, otherwise one runs the risk of ending with so-called negotiated nonsense or incredible outcomes. We have seen in Chapter 3 that experiments tended to end up with innovations that never reached final implementation because of the resulting uncertainty from the limitations of the experiment. We noted in that specific chapter that games likely had to be designed from a design-science perspective and that instead of validity and reliability, crediblity and usability would become more relevant (Klabbers, 2009). The analysis and subsequent framework for game design then put more focus on usability, by providing an elaborate analysis of the context-of-use of gaming simulation. It stated that usability mainly involves gaming simulation's ability to manipulate process volatility by working on four mechanisms.

However, taken into account the initial analysis in Chapter 3 we still must incorporate in the design of games the role validity plays, subjectively, to those involved in innovating in the railway sector. We feel this strongly relates to the quality of credibility of a tool such as gaming. Then we state for games useful impact on innovation processes, the perceived validity by those involved in translating the DIS to the DIL (via the mechanisms) becomes a crucial factor. In this chapter an collaborative assessment of validity is proposed for the debriefing. Credibility is then not seen as directly playing a role in gaming's value for systemic innovation processes, but rather functions as a catalyst for games' usability to become more pronounced.

In addition gaming simulation sessions occur in relative isolation from the context in which it is embedded: games occur parallel to contextual processes (see Klabbers, 2009 model on micro- and macro-cycles in Chapter 3). However for games to have a true value they need to influence the mechanisms at play in the context. Some of these influences might occur without the active input from a game designer or facilitator as games are not entirely in isolation, think of game players and observers returning to their real life work environment. We however assert here that the debriefing is a suitable environment were such influences can be 'designed'. It is in the debriefing that innovations can be more fully disentangled, interlocking opportunities with other projects can be envisioned or shielding can be diminished. Debriefing is the place where the micro-cycle of gaming play can become truly embedded in the macro-cycle of its context-of-use (see Klabbers, 2009 and Klabbers, 2018). This chapter therefore provides a framework to design carefully the debriefing phase of gaming simulation. For a more in-depth analysis we refer the reader to Lo et al. (2013) and Van den Hoogen et al. (2014a; 2014b; 2016). The framework we present in this chapter is a distillation of these works.

9.1 Debriefing

The topic of debriefing is not solely of interest to gaming simulation researchers as the activity merely refers to a collective discussion of events that happened before the debriefing. As such, debriefing is used in many more instances, such as after military operations, traumatic events or after deceptive psychological experiments (Lederman, 1992). Hence, a general definition of debriefing is: “the process in which people who have had an experience are led through a purposive discussion of that experience” (Lederman, 1992: 146). Whereas debriefing is clearly distinguishable from the real-life events that took place before it, debriefing in gaming simulation is much more an intrinsic part of it. For debriefing of educational gaming simulations in particular, learning comes from the debriefing rather than from the game itself (Crookall, 2010). Multiple scholars have pointed to the crucial importance of debriefing in realizing the overall value of gaming simulation, also known as simulation games or serious games. We attribute this to Lederman (1992), Crookall (2010) and Decker et al. (2013). Games are devices that allow experiential learning to be practiced. However, effective learning only comes with reflection (Decker et al., 2013). Debriefing allows for the integration of experience and reflection. It is then noteworthy that little attention is paid to this crucial element of gaming simulation even though scholars have consistently called for more attention (Lederman, 1992; Dennehy et al., 1998; Fanning & Gaba, 2007; Crookall, 2010). We note here that just as we design games differently depending on whether we intend to use them for learning, policy making or research, the way we debrief should also be in line with the game’s purpose (Peters & Vissers, 2004).

We build on previous work that remained on a rather abstract level in order to provide a complete framework for the debriefing of research games (Lederman & Stewart, 1987; Van Ments, 1983). Although we have stated that games’ usability for systemic innovation depend on other factors than those usually related to research games, their credibility is related to the extent it resembles a research game. We then present a framework in this chapter by combining insights from existing literature on debriefing, empirical work on the context in which our gaming simulations are applied, as well as our experience in conducting and debriefing gaming simulations. The framework has both a structural and a methodological component and we provide a topic guide that shows which topics a debriefing should touch upon and a methodology by which these topics should be approached.

In the next section, we first provide a literature review on debriefing functions and phases, in which we present an initial outline of the debriefing structure. This section relies heavily on the existing literature on games for learning, since most work on debriefing gaming simulations has focused on this specific game type. The following section focuses on three gaps we have found in the current literature and serves as the impetus for providing a new debriefing framework for research and design games. First, the framework incorporates the context in which gaming simulation is applied to study innovations and hence becomes part of ongoing innovation processes. Second, validity is a key construct to assess the quality of a research game. The construct can also be found in existing debriefing frameworks for learning games, but its operationalization is not yet sufficient. Validity is a multi-dimensional construct and the framework presented here does justice to this property. Third, the framework provides a methodology by which debriefing professionals can open up the so-called black box of the simulation run.

9.2 Gaming simulation for research

To reiterate the findings from chapters 3 and 8, the complexity of the system we wish to model and simulate leads to many validity issues for which the game itself cannot control. We have proposed that debriefing plays a significant role in adjusting for the flaws of gaming simulation as a research tool but also to capitalize on the promises of the tool to support innovation processes. In addition, the same gaming simulation might perform different functions depending on the observer. For one organizational entity it might serve as a way to rigorously test hypotheses, while for another entity the game provides an ideal opportunity to observe a system holistically and perform a diagnosis. For operators, the game is a way to gain influence in the innovation process or a threat to their autonomy. All of these conflicting expectations potentially have two detrimental consequences: First, the research game is not able to answer one specific research question and converge on a single final design in a way it is perceived to be valid to relevant stakeholders outside of the direct gaming run. Second, the game might create negative effects outside of the game. For both purposes we feel that a debriefing is a valuable, even necessary, addition to the design of a research game.

9.2.1 The role of debriefing in gaming simulation

In general, debriefing is the collective assessment of in-game events and the discussion with game participants about the events' relation to the real world. While such processes are highly valuable for learning purposes, as we contest, the same mechanism will also improve gaming simulation for innovation processes. Even without the need to allow for game player learning, the assessment of in-game events, their significance and their relation to the outside world are enormously relevant for research and design games as a methodology. Consequently, we base a considerable part of the theoretical background for our framework on existing work on debriefing games for learning. The notion that debriefing should be an intrinsic component of gaming simulation design is supported by the fact that experiential learning is a matter of experiencing an event and reflecting on this experience. While the gaming simulation is designed in such a way to provide the player with a realistic experience, the debriefing allows for reflection. As such, many debriefing frameworks for gaming simulation have focused solely on games for learning and applied Kolb's cyclical model of experiential learning (Kolb, 2014) as the foundation (Decker et al., 2013; Dennehy et al., 1998; Van der Meij et al., 2013). This cyclical model portrays experiential learning as moving from experimentation, via experiencing and reflection, to conceptualization. Hence, debriefing usually involves two parts: a collective assessment of what has happened and a discussion on the implications of these events outside of the game (see e.g. Kriz, 2003). Debriefing ensures better task performance and allows players to learn more about a decision domain and develop heuristics to significantly reduce the time between observation and decision-making (Qudrat-Ullah, 2007).

9.2.2 Topics of a debriefing

Most frameworks focus mainly on the phases, or topics, that a debriefing should have. In the realms of games for learning, Sims (2002), Thiagarajan (1992), and Lederman (1992) provide insightful frameworks. However, frameworks for games for research an design, the topic of this chapter, are less developed.

Peters and Vissers (2004) are among the few gaming simulation scholars who specifically target the debriefing of research games. According to them, debriefing of research games has three functions:

1. Providing an opportunity for participants to cool down
2. Protecting the instrument of gaming simulation
3. Validating the researcher's interpretation of simulation outcomes.

At first sight, validation seems the most obvious of the three phases of debriefing. Gaming simulations are artificial environments in which the simulation is open due to the involvement of human game players. This creates internal and external validity issues. The researcher's interpretation of the simulation outcomes should therefore be validated using feedback from participants in the gaming simulation. However, the first two functions are also important. First, we strive for high levels of immersion when we want game players to portray realistic behavior in a game. Thus we ask game players to enter into a reactive mode, dealing solely with the decisions presented to them by the game model and not reflecting on the model itself. In a debriefing we ask that they reflect on what happened, and potentially also ask them to question the model. The transition between these two modes, from reactive to reflexive, does not happen automatically. A cooling-down phase therefore facilitates this. Second, game players are usually scarce, especially in organizational settings where game players are also employees responsible for day-to-day operations. Successive participation or participation by their colleagues is key. We also need to ensure that controversial issues, such as contested innovations tested in the game or conflicts between game players, stay within the realms of the game. As games for research do not primarily look for interventionist effects, what happens in the game should not have any immediate impact outside of the game. The debriefing is the ideal means of controlling this.

9.2.3 A systems perspective on debriefing

Kriz (2010) was one of the first to apply a systems perspective to the debriefing process as games are intended to say something about referent systems or designs in the large. The systems perspective pervades the framework as it acknowledges complex features of systems, its multi-interpretability and path dependence. To do justice to these properties of both the referent system and the game model, Kriz (2010) recommends using six distinct phases in the debriefing process. A key component of this debriefing framework is that gaming simulations allow for the collective and holistic study of complex systems. This collectiveness and holism requires researchers to combine many insights from players and observers and to converge these towards valid propositions concerning the main causal mechanisms that drive the simulation outcomes. An overview of these phases is provided in Table 9.1.

Table 9.1 Phases in a debriefing (Kriz, 2010).

Phase	Topic	Explanation
1	How did you feel?	Cooling down of the participants
2	What happened?	Data collection
3	How are the game and reality connected?	External validity
4	What did you/we learn?	Reaching conclusions
5	What would happen if…?	Testing replicability/sensitivity
6	How do we proceed from here?	Planning for action

Although not specifically targeted at games for research, this framework provides a good direction for the debriefing of research games that involve the study of complex systems. In summary, a properly structured debriefing should contain distinct phases: cooling down, data collection, validity and reliability analysis, planning for action and protecting the instrument.

9.2.4 Missing links

To summarize the literature, frameworks for debriefing seem well developed for gaming simulation for learning, whereas debriefing for research games deserves further attention. We have seen three key phenomena that create need for a more fully developed framework. First, the context in which gaming is applied pervades only slightly in the debriefing framework. Second, validity is not operationalized in enough detail to serve as a structuring force on debriefing frameworks suited for games to become credible to those involved. In particular, the fact that such games are not solely about the design of the game but also about the design of the simulation run is overlooked. This creates the need to incorporate matters such as internal validity in the debriefing. Third, there is no clear methodology to tackle the topics. A topic guide alone does not help the debriefing professional to actually assess all the topics, and instead only points to those topics that require further attention.

9.3 Framework for debriefing games for research

Based on our analyses in the previous chapters, we have developed a debriefing framework. We provide the conclusions we have drawn and the lessons learned over the years. We were involved in designing the ad hoc low-tech tabletop gaming simulations for the Dutch railway sector as mentioned throughout this thesis. Being designed on an ad hoc basis, there were significant differences in the specific research questions addressed. However, a common factor in all the games was the simulation of operational processes of railway systems (trains running according to a schedule, a realistic depiction of the infrastructure, and operators dealing with scenarios such as major disruptions around railway stations). An example of a typical question on which our games were intended to shed light is: does the punctuality of train traffic around the central node of the network increase if we separate two heavily used corridors by removing railway switches? The fact that these questions involved a unit of analysis at system level, and contained both technical and social elements, created the need to use gaming simulation to test such measures in a safe environment.

The framework tackles precisely those problems we have found in the current literature on debriefing when applied solely to gaming simulations intended to test hypotheses. First, it takes into account contextual influences on the ability of a game to test a hypothesis solely by running a simulation (and disregarding the debriefing). Gaming simulation seems to have a, sometimes undesirable, natural tendency to allow for exploration rather than explanation, caused by contextual influences. Our debriefing helps to counter this tendency. Second, it uses a topic guide that operationalizes validity in more detail. Thus our framework helps to alleviate many of the inherent validity threats of using a method that lingers between field observations and classical experiments. Although, as we have stated, validity in itself is not a necessary quality of gaming simulation in innovation processes, its perceived validity (or credibility) however is. Tackling credibility issues in the debriefing of a gaming simulation then will highly resemble any other validity assessment of a research instrument. Third, it provides a specific methodology that enables the debriefing to open up the 'black box' of the simulation run. This last contribution is significant in that a topic guide alone barely helps in actually debriefing a research game. For instance, a topic guide tells you to assess ecological validity, but does not provide you with a method for actually doing so.

In addition, our framework truly coalesces gaming and debriefing to make it an intrinsic part of the discipline of gaming simulation. This is because the gaming simulation and the debriefing mutually reinforce each other in both ways. Using our framework allows the debriefing to alleviate many of the intrinsic validity and reliability threats to gaming simulation, threats we have already elaborated on in Chapter 3. By having the debriefing focusing partly on tackling these threats, other qualities of gaming simulation can be more freely designed in the game design process and the simulation run. This is because as we have stated in Chapter 3, validity and usability are sometimes at odds with each other. Conversely, a carefully designed gaming simulation helps to improve the debriefing. In the end, the use of our framework improves the entirety of the game and the debriefing: they become a whole.

9.3.1 Context

The gaming simulations we have designed, employed and debriefed up to now were used as applied experiments in which organizations could explore or test innovations [See Meijer (2012) and Lo et al. (2013) for an overview]. Hence, gaming simulation is not an isolated phenomenon, but is embedded in ongoing technological, social and institutional processes over time. The design of a gaming simulation needs input from the environment, e.g. innovations to be tested, data, models, game players, and feeding back of results of the simulation. Two parameters seem particularly relevant in this case for both the innovation and the function of a game (Van den Hoogen and Meijer, 2015):

1. Innovation processes can be either *stable* or *volatile*, or move from one to the other over time. Volatility entails rapid changes in the design of the innovation, rapid entrance and exits of designers and decision makers and fluid and flexible institutions that govern these activities.
2. Gaming simulation can either create *convergence* or *divergence.* Divergence is the exploration of a multitude of designs, the opening up of the arena of designers and the exploration of viable institutional arrangements. Convergence is the opposite, where designers and decision makers become more fixed, increasingly focusing on a single design as the final option under increasingly stable institutional arrangements.

This conceptualization of the value of gaming simulation in light of the context in which the method is applied led us to further study the practical value of using so-called explanatory research and design games. We term these explanatory for their intended ability to have a diverse range of stakeholders gradually focus on one design and one process by which to implement it: the so-called convergence or contracting of the P, S and I-spaces.

An explanatory game diminishes volatility because the design of the game and the simulation run is such that it should allow stakeholders observing the game to focus on convergence to one final design solution. Because of this function of gaming simulation, we have seen that the method is often employed in times when the innovation process is highly volatile. In these times, when many designs, ideas and innovations float through the organization and many new designers, decision makers and other stakeholders enter the decision-making arena, stakeholders view gaming simulation as a proper tool to alleviate this volatility. Then, given this purpose validity assessment become more important. In chapter 3 we have already elaborated on the relevance of perceived validity for the legitimacy and credibility of the method of gaming to relevant stakeholders.

However, contextual influences of this volatility have a pervasive effect on the ability of gaming simulation to actually create convergence (see also Van den Hoogen and Meijer, 2015). Then, gaming performs poorly. Firstly, in highly volatile times, rapid changes occur in the set of feasible design alternatives. On multiple occasions, we had to make last-minute changes to the game model to incorporate alterations in the innovation or in other relevant parameters. Since there is always a lag between the input for the game design, the design of the experiment and the output in the form of results, there is a chance that the game answers questions that are no longer deemed relevant by the organization. Such feature of the use of gaming in policy-making processes is already elaborated on by Klabbers (2009). Secondly, and much more significant if one intends to create convergence with a gaming simulation, is that gaming simulation serves as a window of opportunity to test other innovations as well. Especially in the capital-intensive and safety-critical industries in which we operate, there is little opportunity to test innovations. When organizational entities other than the primary client of the game become aware of the possibility to test their innovation in a gaming environment, we see an influx of additional research questions. The usual way of going about adhering to these questions would be to increase the factorial design of the experiment. However, pure experimental research often demands full factorial designs (making all possible combinations of innovations), resulting in exponential increases in the numbers of runs. Due to time constraints, real-life operators are usually only available as game players for a limited time. The choice is then to either not test them all or to make the simulation more abstract and omit real-time play, thus risking lower levels of game player immersion. These are design choices that endanger internal and external validity, respectively but increase usability for the context-of-use, the broader railway sector, immensely.

Thirdly, immersion is sometimes a problem in itself. Game players are usually operators who enjoy a certain degree of autonomy in their daily work. This creates both a desirable and undesirable distance between those who carry out the work (and are part of the game) and those who design the overall system in which the operators are placed (and observe the game). When this distance is removed by employing a gaming simulation, two phenomena can occur: game players either feel under heightened scrutiny and start behaving differently than in real life, or they feel heard and desire a dialogue with the designers of the innovation during the game. Both phenomena create immersion issues since we want game players to behave just as they would in real life where there are no designers observing or able to communicate with them. Klabbers (2009) already pointed to the problems of self-referential nature of gaming elements such as human actors and social organizations. By classifying, as the outside observer, those players as seemingly relevant to participate in a game, they become aware of their relevance, more so than in real life.

Making the debriefing an intrinsic part of the game design helps significantly in alleviating the aforementioned context-based problems. First, the debriefing can serve as a way of testing all innovations while keeping the number of runs relatively low. This is valuable since a low number of runs enables the game designers to use real-time play, a design parameter often, but not always, associated with high levels of immersion. In the design of the simulation run, game designers and the innovation managers involved can decide together which innovations truly need to be tested in the game run and which innovations can be assessed in the debriefing. Also, properly taking into account the possibilities of assessing additional innovations in the debriefing helps make the gaming simulation more adaptable to last-minute changes in the innovation. To achieve this, a robustness analysis should be included in the debriefing. Game players and observers can concertedly assess the extent to which simulation outcomes will differ if either the innovation changes later on or additional innovations are introduced. This is an important part of the debriefing since one cannot expect the innovation tested to be exactly the same when it is implemented, especially in volatile times. Furthermore, the debriefing can be used to postpone the inherent tendency of gaming simulation to lead to a dialogue between game players and observers. This means that incorporating the debriefing allows the facilitator to better manage expectations. The facilitator could demand that dialogue be non-existent during the game run, thereby increasing immersion and subsequently credibility of the results to outside observers, and in return promise that the desired dialogue will take place during the debriefing.

9.3.2 Substance

Based on the assumption that innovation stakeholders in the railway sector operationalize credibility of gaming simulation outcomes via analytical science concepts such as validity and reliability, we use these latter concepts for the substance of the debriefing. As mentioned in Chapter 3, these concepts stem from the analytical sciences and are strongly related to hypothesis-testing research. Such research is in essence an experiment in which one or more independent variables are manipulated to investigate their effects on a dependent variable (Zechmeister et al., 2001). Two streams can be identified for experimental research: the first is a classical linear perspective on causality and the second is a complexity perspective. The classical linear perspective sees experimental objects as trivial machines, which implies that the same treatment given to the same participant will always have a similar outcome. The complexity perspective takes non-triviality into account, which implies that systems with dynamic feedback show path-dependent and chaotic behavior. In line with this perspective, units of analyses are therefore respectively regarded as black boxes or as a collection of interacting elements. However, two critical concepts are key to the determination of the quality of both streams of experimental research: reliability and validity (Lo et al., 2013).

9.3.2.1 Reliability

Measurement reliability

Measurement reliability is the extent to which a research method or measurement tool provides a similar value if the measurement is repeated (Messick, 1975). In quantitative terms, the reliability of the measurement tool can be expressed as a margin of error. For instance, if a thermometer should be measuring a temperature of 39 degrees Celsius, but indicates a value of 38 half of the time and a value of 40 degrees for the remaining measurements, the margin of error of the measurement tool is 1/39.

Sensitivity

The sensitivity of the experiment is often determined in computer simulation experiments, in which the researcher determines whether similar causal relationships are found when the experiment is repeated with exactly the same sample and setup. This complexity perspective on reliability follows from experiments with dynamic feedback systems. Because dynamic feedback systems inherit stochastic and sometimes chaotic properties, different results can be found when experiments are repeated with the same or almost the same starting conditions. An indication of the sensitivity of an experiment is useful in order to assess whether the results are sensitive to the initial conditions or to critical decisions by game players.

9.3.2.2 Validity

Internal, external and measurement validity are the core validity types in experimental research, in which external and internal validity play a dominant role in determining the quality of the experiment (Zechmeister et al., 2001).

Internal validity

In establishing a causal relationship, the research needs to meet the conditions of co-variation, time-order relationships and elimination of plausible alternative causes (Zechmeister et al., 2001). Co-variation is the first step in establishing a causal inference, which can be fulfilled by finding a relationship between the independent and dependent variable. In identifying the cause and effect for the independent and dependent variable, a time-order relationship can be established. Lastly, confounding variables need to be isolated to eliminate plausible alternative causes.

External validity

External validity has multiple definitions that are subject to conflicting interpretations (Morton & Williams, 2010). We distinguish external validity in terms of generalizability, i.e. results that can be transferred from the current sample to the population, versus ecological validity, from the simulated environment to a real-world setting, which would be in line with the 'fieldness' of the experiment (Harrison & List, 2004). Selection of a representative sample ensures the generalizability of the results as a reflection of the population. Parallel resemblances can be drawn for ecological validity with the three gaming simulation validity types defined by Raser (1969). Gaming simulation validity can be broken down into structural validity, process validity and psychological reality. The simulated gaming model may be rather abstract or simplified in terms of processes, interactions, and contextual and physical cues in comparison to the reference system. As such, the omitted characteristics of the reference system may endanger the transfer of causal claims made within the gaming simulation to the real world. Applying a sensitivity analysis could support the assessment of this type of external validity by focusing on whether parameter sensitivity, tipping points and critical decisions by game players could be a resemblance of events in the reference system.

Measurement validity

Measurement validity, also known as test validity, refers to the validity of the measurement tool or instrument itself. Psychometric researchers have predominantly focused on the different typologies involved in the use of measurement instruments, often questionnaires. The American Psychological Association, American Educational Research Association, and National Council on Measurement in Education have set Joint Standards (Campbell and Stanley, 2015). Construct, criterion and content validity are distinguished as the three main measurement validity categories.

9.3.2.3 Topic guide

Based on the literature review in the previous sections, we have identified eight phases that need to be addressed in a debriefing session of an explanatory game, in which a large overlap exists with existing literature by Kriz (2010) and Peters and Vissers (2004). However, this chapter recognizes the gap in the existing literature regarding the specific topics that need to be addressed within the validity and reliability analysis phase as these concepts determine to great extent the credibility of gaming simulation' outcomes. To also incorporate the context of volatile innovation processes, the topic guide introduces a robustness analysis, determining to what extent the outcomes are robust against slight changes in the innovation. Table 9.2 summarizes the findings from the previous sections and integrates the different debriefing phases with the topics and the ideal participants involved for each phase.

9.4 Method

Through the use of gaming simulation we are interested in testing and designing innovation and determining on the socio-institutional measures to implement such innovations. Markedly different from classical medical and psychological experiments, however, is the fact that we apply the innovation to a model of a reference system, which is the game model, rather than to a single atomistic entity. This system comprises many interdependent elements in a web of complex causal relationships and adaptable human game players. We explain this difference using the notion of trivial machines (TMs) and non-trivial machines (NTMs) by Von Foerster (1984). These notions subsequently impact how we can claim any causality after experimenting with systems and designing future innovations in such systems. Barreteau et al. (2001) and Klabbers (2009) have pointed to this feature of modeling and gaming simulation and the necessary role of opening the black box for any validity or credibility assessment to outside stakeholders. Such opening of the black box of the game run, and systematically assessing what has actually happened during the game, therefore helps to find credible leads for causality in NTMs such as our gaming simulations. This increases the extent to which the usability of gaming, through the mechanisms we pointed out in Chapter 8, becomes manifest.

Table 9.2 Framework for debriefing a research game consisting of phases, topics addressed and participant involvement.

Phase	Description	Topics	Participant Involvement
Cooling down	Change game player's mental state from immersion to retrospection.	Experience Emotions	Facilitator Game players
Data collection	Additional qualitative data from players, observers and facilitators.	Measurement reliability Validity	All participants
Reliability	Assess whether repetition would result in similar outcomes.	Sensitivity	Game players Observers
Internal validity	Can we state with confidence that the experienced causal claim holds within the game situation?	Potential confounding variables	Game players Observers
External validity	Assess whether causal claim holds in real life (ecological) and for different samples (generalizability).	Game artificiality Impact of omissions in game model Sample-specific behavior	Game players Observers
Robustness	Do variations of the tested innovation, or the introduction of additional innovations, create strikingly different outcomes?	Longevity of the relevance of outcomes if innovation processes persist	Game players Observers
Planning for action	Determine what follow-up questions need to be answered. Determine what concrete actions need to be taken and by whom.	Future research questions and actions	All participants
Protect the instrument	Evaluate gaming simulation session. Determine what outcomes may be shared. Ensure a durable relationship with game players.	Experience Emotions	Facilitator Game players

9.4.1 Trivial and non-trivial machines

In traditional experiments, researchers assume that some conceptual device transforms the input *x* into output *y* and that this transformation is both linear and independent of context, time and history. In those instances when the relationship between *x* and *y* is established and the researcher is solely interested in prediction, there is no need to open up the black box of this device. How *x* causes *y* is irrelevant. In contrast to these trivial machines, non-trivial machines bring about causality in a far more complex manner. NTMs are devices in which the transformation of *x* into *y* is highly dependent on history, time and context and in which the device itself changes as a result of *x*. Social systems, consisting of adaptable and interdependent human beings, are perfect examples of NTMs (Klabbers, 2006). Here, how *x* is transformed into *y,* becomes highly relevant and the researcher thus needs to open up the black box (Von Foerster, 1984). Because we assume that the systems we manipulate in a gaming environment are like NTMs, we cannot simply perform a pretest and posttest with and without an innovation, as is customary in classical psychological and medical experiments.

9.4.2 An ontology of events and processes

Researching non-trivial machines in which causality is brought about by an interplay of complexity, path dependence, chaos and interdependence on multiple levels of analysis is common in the more qualitatively oriented fields of the historical and sociological sciences (Griffin, 1993; Hedström & Bearman, 2009). Here researchers rely heavily on narrative explanations that allow them to better describe what is actually going on and also to better incorporate the highly relevant context. This explanation is based on event sequences rather than relationships of variables (Abbott, 2001; Geels, 2011). Hence, an example of a usual description is that the Great Depression in the 1930s, Event A, partly triggered the Second World War, Event B. According to Weber (1949), most events are too complex to state any causal generalization about them. So claiming that economic decline and the likelihood of war are always causally related becomes infeasible. In addition, in contrast to linear causal models, narratives allow the researcher to gain insight into the complex interplay between social structure and human agency over time (Giddens, 1979; Griffin, 1993; Sewell 1992). Narrative style explanations also gained more popularity in the fields of management sciences, as topics became ontologically more complex and linear models failed to acknowledge this. Examples of this can be found in innovation management and organizational theory research (Langley, 2007; Pettigrew, 1992; Tsoukas & Hatch, 2001; Poole et al., 2000) and in research on transitions of sociotechnical systems (Geels, 2011). Since what happens in a gaming simulation is really a sequence of events rather than a link of variables, their methodologies could support our debriefing. Our games are more discrete-event simulations than system dynamics models and observing them thus needs to acknowledge the 'eventness' of the simulation.

9.4.3 Methodologies

Of all the methodologies applied by historians and sociologists, event-structure analysis seems to be the most developed (Heise, 1989). Event-structure analysis enables the researcher to structure events and portray how accumulations of past actions constrain or instigate future events. For a better overview of these methodologies we refer to Manzo (2010). The event-structure analysis starts by drawing a timeline of the events that have occurred. In other words, the events have a specific temporal ordering. Next, one must determine the extent to which an event causally triggered the next event or another event later on. Key elements of this assessment are counterfactuals, which are negations or modifications of a specific event and basically involve asking 'what if' questions (Griffin, 1993). If Event A1 occurred, could Event A2 also have occurred? The third step is to determine whether these counterfactuals are objective possibilities (Weber, 1949). This means that the counterfactual is in itself realistic and remains conceptually close to the real past. If the hypothetical negation or modification of the event would have caused a completely different unfolding of events later on, this event is a causal triggering for all subsequent events (Griffin, 1993). To assess this counterfactual world, researchers can either use other cases as a benchmark or theoretically deduce how the story would unfold. In Table 9.3, we briefly summarize the steps commonly found in narrative analyses that focus on causality.

Table 9.3 Event-Structure Analysis, based on Griffin (1993).

Step	Action	Description
1	Determine events	Map all game player decisions, changes in game parameters and context.
2	Determine counterfactuals	Map for the potential counterfactual events for every event.
3	Assess realism of counterfactual	Determine whether the counterfactual is close to the real past and is realistic in real life.
4	Determine counterfactual world	Assess to what extent the different event would trigger different subsequent events.

Using this methodology helps to tackle all the topics in the aforementioned topic guide in a more systematic manner. The collective determination of events improves the data collection phase of the debriefing. In the years we have spent designing gaming simulations for the railway sector, we have found the tool to be an ideal method for enabling multiple stakeholders to holistically observe processes that would otherwise be separated in space and time. In addition to the more quantitative data usually logged during gameplay, more qualitative observations are possible. Qualitative data is valuable for two reasons: first, it is better able to capture the complex nature of the dynamics that occur during gameplay. Second, it requires less operationalization beforehand. This increases the possibility of testing innovations for which the performance measures are still being debated or hard to quantify. To fully benefit from this in the debriefing, there must be considerable attention for data collection during the design of the game and the experiment. Observers, most often designers of the innovation and subject matter experts, need to be present during the game and given instructions. Whereas retroactive accounts of game players cannot be identified beforehand, it is possible to determine what observers should look for in advance. For instance, observers could be provided with a topic guide. During the debriefing, all observations are shared to form a common picture of what has happened during the game. This serves both the purpose of calibrating the observations and improving the measurement reliability, and of concertedly creating a chain of crucial events.

Now, in addition to how variables changed during gameplay, we are discussing the event chains that caused these dynamics in the variables. In other words, we are opening up the black box. In this phase of the debriefing, we map all key events that occurred during the game. We use the musical staff as a metaphor, with each line representing an element of the system, for instance: game player, train and station. The notes are events instigated by the element. The story is the temporal progression of events. For instance, a train might break down as Event A, which invokes a response by a traffic controller as Event B, and so on.

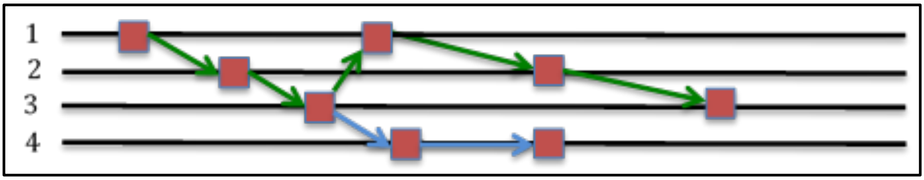


Figure 9.1. A Four-Element Event Chain System. (Van den Hoogen et al., 2014: 3511).

Figure 9.1 presents a graphical example of this, with four elements and the green path representing the actual events and the blue path representing the possible alternate decisions. However, it should be noted that the level of detail we use here is merely for didactic purposes. The level of detail we usually apply is much lower, focusing on around 10 events that best describe the gameplay. Some of the questions we use to draw up such an event chain are: What happened? What was crucial for the experienced gameplay? What processes did you observe?. The role of the facilitator is to combine all of these insights, assess their congruence and juxtapose contradictory observations.

The event chain becomes especially valuable for the systematic assessment of the validity and reliability of gaming simulation outcomes. In addition, it provides a good method to collectively discuss the impacts of inherent internal and external validity issues. First, the internal validity of the causal claim is increased by determining *how* the innovation brought about changes in variables rather than simply stating *that* the innovation did so (George & Bennett, 2005). Second, if we want to determine to severity of validity threats, both internal and external, we can use the event chain analysis to assess whether simulation outcomes are highly dependent on certain validity-threatening phenomena. For instance, we can assess whether certain events are triggered by omissions in the game model (to test its ecological validity), whether one decision by a game player could have just been another decision resulting in a different unfolding of events (to test its sensitivity), or whether other game players who were not involved in the game would have decided something different for a certain event (to test its generalizability).

The event-structure analysis would enable the facilitator to study the sensitivity of the simulation outcomes to validity issues. Critical decisions can be assessed to determine whether the player could have just as well have decided something else, and to what extent this would have caused a completely different unfolding of events, as represented by the blue path in Figure 1. Both game players and observers usually determine which decisions were critical. Immersed game players usually cannot recall all the events that took place during a game, but do know the significance of the decisions they have made. Observers are less aware of the significance but are more likely to recall decisions, especially when are tasked to do so. For this collective imagining of a different unfolding of events, we rely on a mental simulation of the changed game. An advantage of this is that low-tech gaming simulations are easy to re-use and hence can serve to support this analysis. If the mental simulation places too much cognitive strain on the game players, we can use the game that is still available to quickly replay a few events.

9.5 Synthesis

In Table 9.4 we present a brief overview of a possible debriefing of a research game. For a more in-depth look at how we applied part of this framework in specific cases in the Dutch railway sector, we refer to Van den Hoogen et al. (2014b). The framework incorporates existing notions from the literature (Peters & Vissers, 2004; Kriz, 2010) and adds a more thorough operationalization of validity and a methodology by which to actually study validity in the debriefing. In addition, the table shows the ideal roles of each participant.

Table 9.4 Reflection of Validity and Reliability Issues during Debriefing.

Phase	Participant Debriefing Roles		
	Player/Operator	Observer/SME	Facilitator
Cooling down	Taking a break, discussing game experiences	Summarizing observations	Leading discussions on game experience
Data Collection	Establishing event chains	Establishing event chains	Juxtaposing statements; assessing measurement validity and reliability
Sensitivity	Determining counterfactuals and their effects on subsequent events (based on experience)	Determining counterfactuals and their effects on subsequent events (based on theory, rules, etc.)	Asking players and observers about crucial events and objective possibilities
Internal Validity	Determining how treatment impacted the events; determining effect of confounding variables	Determining how treatment impacted the event chain; determining effect of confounding variables	Identifying potential confounding variables due to experimental context
Generalizability	Comparing own decisions with probable decisions made by peers; comparing sensitivity of decisions to changes in other dimensions of the sample: different timetable, etc.	Identifying differences between the sample and the population	Linking differences found by observers with players' comparisons
Ecological Validity	Determining perceived realism and effect of omissions of elements and processes of referent system on event chains	Determining the effect of omissions of processes and structural properties of referent system on event chains in game	Discussing what omissions were applied during game design
Robustness	Determining effects of changes in innovation and introduction of additional innovations on event chains	Determining in what ways the innovation might change later on	Introducing the agreed-upon leftover category of innovations not tested in the game run
Planning for action	Determining to what extent other operators are able to handle the innovation once implemented, and if additional training is needed	Determining what follow-up research is needed and how concrete actions will be coordinated with all stakeholders	Summarizing findings of the previous discussions to start up this phase
Protect the instrument	Discussing to what extent the innovation or the game was controversial and what can and cannot be fed back into real world	Discussing to what extent the innovation or the game was controversial and what can and cannot be fed back into real world	

9.6 Conclusion

This chapter has delivered a framework for debriefing games for research and design. In particular, we zoomed in on games used for innovation. Discussing robustness and planning for action are especially important for this specific context of use, whereas the other phases are valuable to increase the credibility of simulation outcomes to relevant outside observers. By using our framework we tried to tackle the context, substance and method of debriefing games. The debriefing framework enables a gaming simulation to do justice to the volatile context of innovation processes, a context that we explored in previous chapters. By making debriefing intrinsic to the design considerations of a game, the game designer is better able to cope with this volatility. As far as substance is concerned, the framework delves deeper into the specific topics a debriefing should address: data collection, sensitivity, internal validity, generalizability, ecological validity and robustness. Event-structure analysis, a method used in the qualitative historical and sociological sciences, allows for a more thorough and rigorous analysis of causality and validity, thereby opening the black box of the game run.

Limitations of this framework are twofold. First, we distilled the framework from the many experiences we gained in designing and debriefing a multitude of different games. However, the applicability of the framework in its entirety has yet to be tested. Future research should look at the feasibility of rigorously applying this entire framework in a debriefing. In addition, such a study could examine whether or not the framework improves the gaming simulation, by whatever metric. And for the practitioner, it could result in a set of exemplary questions that operationalize the dimensions and phases mentioned in this chapter. Second, the method we propose requires game players to mentally simulate the answer to 'what if' questions. Given that the focal point is a complex system, the extent to which game players are able to do is still debatable. However, their ability largely determines the validity of the claims we make on basis of the debriefing. Are the results really robust or is the game player simply unable to perceive that a slight change in the innovation will bring about radical changes in the dynamics of the system? Future research could look at their cognitive capacities as well as methods to improve the collective assessment of alternate courses of gameplay. Nevertheless, we feel that debriefing is an intrinsic part of designing games for research and design as well as games for learning. We have seen how debriefing has become more and more intertwined with the designing of models and simulations. With this framework, we intend to improve this cross-fertilization between gaming and debriefing.

10 Conclusions and Limitations

In this final chapter we answer our research question. Our research question involved three levels of analysis: macro-level patterns of innovation processes, driving mechanisms, and the working of games on patterns through these mechanisms. In the introductory chapter we put forward the question:

"What mechanisms play a role in driving systemic innovation process in the Dutch railway sector and in what ways is gaming simulation able to influence relevant macro-level patterns through these mechanisms?"

After considering the previous chapters we can state that through the careful design of the game, the game design process and especially the debriefing, games enable innovation stakeholders to more intelligently control four relevant driving mechanisms in order to manage process volatility. We found that process volatility was the key macro-level pattern that set apart systemic innovations from other innovations, that this pattern was both problematic and functional, and that four mechanisms drove this specific pattern. Games relation with these mechanisms is profound but also complex. Rather than designing games for innovation as if they were classical experiments, stakeholders need to design games in such a way that they allow for the attenuation or alleviation of the relevant driving mechanisms. This deviates strongly from initial conceptions, from an scientific-analytical perspective, held by us as researchers as well as by the railway sector itself. It also strongly impacts the way games should be designed.

10.1 The fuzzy back-end

Based on the multiple case study of Chapter 6 we conclude that the gradual increase in volatility is what sets systemic innovation processes apart from other innovation processes. This volatility encompasses dynamics in what the innovation constitutes, who is involved in implementing it, and what institutions stakeholders use to govern the process. Using a pre-structured analytical framework we studied three different innovation processes. These processes involved the introduction of a traffic management system in the UK, the building of a railway tunnel and the introduction of Japanese design principles in the Dutch railway system. The latter being systemic, we found that for these processes the volatility was located at the final stages of the process, rather than at the beginning.

Whereas linear and multi-level perspectives on innovation assume convergence in these dimensions of the process, we see that systemic innovation processes diverge from initial stability to volatility later on. Rather than a fuzzy front-end, these processes have a fuzzy back-end. Only in later stages of the process did we see many changes in the innovation artifact, the arena of involved stakeholders and the applicable institutions.

Chapter 6 took on a solely structural perspective, disregarding individual actor behavior and strategies. We saw that the innovation's loosely-coupled nature and its inability to build up momentum caused the gradual increase in volatility over time. Systemic innovations, being sets of otherwise more incremental changes, were more influenced by interactions with the regime and with other ongoing innovations than innovations that increasingly became fixed. Since the individual elements of the innovation were mostly incremental changes, the innovation did not need much upfront material, financial and organizational support. Hence, interactions with the regime mainly took place later on in the process as the nature of the innovation permitted the initial stakeholders to remain under the radar for longer periods of time. We posit that for this reason systemic innovation processes need very different management principles than innovation processes that inherently see increasing stability.

10.2 Four driving mechanisms

From our analysis on a single case study in Chapter 7, we can conclude that four relevant driving mechanisms underlie the macro-level pattern of increasing volatility. We studied the transition of the Dutch railway system towards a system designed according to Japanese principles. We analyzed the process of the innovation from an idea in 1997 to partial implementation during the Utrecht Central Station renovation project. We found that *ambiguity*, *disentanglement, interlocking,* and *shielding* were the most relevant mechanisms in moving forward the innovation process. Uncertainty reduction or knowledge creation was not found to be a relevant driving mechanism. In addition, these mechanisms alone and in their interaction with each other explained the volatility pattern we found earlier. These mechanisms appear when the process is studied not solely from a structural perspective but also from the perspective of agency, i.e. how actors actually deal with the challenges of implementing a loose set of innovation elements, which in conjunction can be considered systemic. The four mechanisms are:

Ambiguity
The process of maintaining an image of an innovation in such a way that inherent dilemmas remain invisible and most stakeholders perceive it as a win-win situation. Ambiguity serves to keep an innovation as an idea alive whilst not yet being implemented.

Disentanglement
The process of operationalizing a systemic innovation. From a set of abstract notions on what might constitute this innovation to an operationalized innovation. During this process the innovation becomes less systemic.

Interlocking
The latching-on of one innovation onto another innovation. Interlocking might serve two purposes: firstly, it allows an innovation to use a window-of-opportunity and makes the

innovation more implementable. Secondly, interlocking is an effective way to cope with complexity.

Shielding
The process of keeping away regime pressures and pressures of concurrent innovation processes from ones' own innovation process. Shielding enables stakeholders to design the innovation beyond the scope of their responsibility.

We saw in Chapter 7 that these mechanisms contain a constructive side: by invoking these mechanisms innovation actors are able to advance an innovation process. They also have a disruptive side: over time these mechanisms tend to push inherent dilemmas away in time and space. This influence on where, how and when dilemmas pop up and how they can be dealt with is the reason for volatility being located at the fuzzy back-end of systemic innovation processes.

From our single case study we can also conclude that the four mechanisms interact with each other. This interaction is non-trivial as only in their conjunction could a systemic innovation be implemented. Disentangling an innovation causes the innovation to loose parts of its systemic properties. However, if innovation actors interlock the innovation with other innovations while shielding their own innovation from otherwise increased interactions between these innovations, they are able to implement a disentangled innovation that will have systemic effects later on. We saw at DSSU that the innovation idea started as systemic, ended as systemic, but was less systemic midway in the process when it interlocked with the renovation of Utrecht. Our case study showed that this loss of systemicity was functional, it allowed parts of the innovation to interlock with the renovation project, and that the other three mechanisms played a role in making the innovation more systemic after the first phase of implementation.

The analysis on the phases after the interlocking with the Utrecht renovation project showed that these interactions also cause mechanisms to shift from constructive to disruptive. For instance, under high levels of shielding, innovation actors tend to interlock their innovation with other innovations but solely on a technological level. They do this to deal with the complexity they encounter. Innovations then do not interlock on actor and institutional level. This created the effects that many innovations became interdependent without the interdependence of innovation stakeholders. When other innovations changed, the impacts of this on the design of Utrecht were not taken into account.

Combining insights from Chapters 6 and 7 we arrive at two relevant conclusions for the use of gaming: firstly, the macro-level pattern of increasing volatility is both problematic and functional. All four driving mechanisms played a role in allowing a systemic innovation process to be moved forward and while doing so kept volatility at low levels. This came however to the detriment of mechanisms persisting and becoming disruptive in later stages, hence increasing volatility. Secondly, the macro-level pattern was highly emergent. The gradual increase in volatility was not the result of intentional actions by stakeholders. Rather this pattern was the result of many interactions over time where no one involved in the process intended to create this overall pattern. Given that volatility is the result of a dilemma between progress and coordination and that the network of relevant stakeholders were not able to deliberately deal with this dilemma, we posited that this is the role of gaming simulation.

10.3 Games, mechanisms, and patterns

A case study on the use of gaming simulation showed that games have effects on volatility and that this effect is mediated by the four mechanisms. Game design parameters, related to games being either exploratory or explanatory, correspond to respectively front-loading or suppressing volatility. Exploratory games are able to increase volatility before it would otherwise occur. This front-loading effect happens through the open nature of games: they allow for many innovations to be discussed, new stakeholders to enter the arena as well as new institutions to be explored. Explanatory games are games that allow innovation actors to narrow down on single solutions, a single set of stakeholders and they enable stakeholders to choose a single set of applicable institutions. Through this, explanatory games suppress volatility.

The case study also resulted in the conclusion that the relation between game design, mechanisms and volatility pattern is complex. We saw that seldom a game is solely exploratory or explanatory. Game design choices, such as openness of the game, the use of real-time play or the granularity of the game model have different influences on the four mechanisms which in turn leads games to become partly exploratory and partly explanatory. In some instances, game design choices, made to make it more explanatory, resulted in more exploration during game play. This is an inherent feature of gaming simulation. For instance, real-time play increases disentanglement but the subsequent need for more accurate data to feed the model will decrease the level of shielding: one needs input of other departments and they subsequently become aware of the innovation and might desire influence. Optimizing games to optimize mechanisms is therefore impossible and designing games inherently involves trade-offs.

10.4 Designing games

Based on our analysis of game design and its relation to mechanisms and patterns we found that this relation is complex and also partly beyond the direct control of the game designer. Games are not niches because outside influences pervade in the realms of game design, game execution and debriefing. These influences are sometimes contradictory. A game might be designed for one stakeholder to converge on one solution while other stakeholders mandate the testing of additional solutions as well. Also games have a dynamic of their own, due to the incorporation of human game players, observers and facilitators. These phenomena impact the ability of games to bring about either the front-loading or suppression of volatility.

The volatility-suppressing characteristic of games is cumbersome. This is because we saw in Chapter 7 that volatile contexts cause games to become volatile as well. In later stages when a process might benefit from decreasing volatility, such a process would need an explanatory game. Rather we see that in those instances, the volatility of the context resonates in the volatility of game design process, game design and game play. These games then have the tendency to become exploratory, which is detrimental to for instance perceived validity and reliability. Because of this, the debriefing becomes the crucial phase of gaming simulation. Only here would one be able to converge on single solutions and actions and collectively assess the validity and reliability of the outcomes, thereby increasing the credibility of the outcomes and increasing the chance relevant mechanisms are affected outside of the game.

We saw that front-loading volatility is less problematic in our case study on the use of gaming. Games alleviate shields, make the innovation less ambiguous, and allow for the exploration of many interlocking possibilities. This comes however at the cost of the ability to disentangle an innovation. Usually games for this purpose tend to result in a range of ideas on what to do, how to it and with whom. To counter this problem, debriefing is the phase of the game to influence all four mechanisms in such a way that volatility is front-loaded but still the results are actionable.

10.5 Debriefing is most important

From our conclusion that it is the way games can manipulate volatility that determines its value and the observation that during game design and game employment this working on volatility is partly beyond the direct control of the game designer, we found that the debriefing is actually the most important phase. Of all elements of gaming simulation (design, play, debriefing) the debriefing is the phase where involved stakeholders can mostly control the impact of games on mechanisms.

Based on our experience debriefing gaming simulations for innovation processes in the railway sector we concluded that a more structured approach is needed. This structure involves substance (which topics?) method (how to do it?) and context (in light of what?). In general, debriefing a gaming simulation involves collectively opening up the black box of a gaming session. One needs a suitable method to do this. Event-structure analysis, from the sociological and historical sciences, is most helpful. With this method, one is able to bring together different insights from game players and observers, and assess the validity and robustness of the outcomes. These qualities are important for credibility and usability respectively. By doing so, debriefing allows for credible and actionable outcomes to flow from the game, whether the game itself was designed for exploration or explanation.

10.6 Implications

The conclusions have implications for both academic research on innovation processes as well as for practice. Regarding the latter, there are implications for the management of innovation processes and the use and design of games in order to support these processes.

10.6.1 Our findings and existing theoretical work

The main theoretical framework regarding the dynamics of innovation processes dealing with transitions and systemic change was the multi-level perspective (MLP). This framework takes on a highly structural approach, mainly looking at how different levels of structuration (from highly inert regimes to experimental niches) create the opportunities for systemic change. Our findings, derived from an analysis that looks not only at structure but also at agency, have several implications for this theoretical framework.

Firstly, our study implicates that technological change and changes in the social and institutional context are highly interrelated and that theoretical frameworks such as the MLP need to incorporate this. MLP mainly focuses on changes in social actor arenas and institutional setups as an innovation moves from niche to regime but disregards the dynamics in the innovation itself over time. We saw for instance how disentangling an innovation made it less systemic but also allowed for a more confined stakeholder arena and that this confined arena again was better able to further disentangle an innovation. We also saw that when an innovation is loosely coupled, the regime is better able to cherry pick elements of it as it saw fit and that this again impacted the design of the innovation further on. We believe that concepts from the NK literature, such as epistasis and modularity, can help to further enrich the MLP. It would then be interesting to study how niche development, vision setting and network building, three key processes in the MLP and Strategic Niche Management literature, are related to changes in innovations such as the gradual coupling or growing modularity of innovation elements.

Regarding these three aforementioned key concepts of the MLP, our findings implicate that these concepts are not beneficial per se nor do these key concepts reinforce each other automatically. We saw for instance that niche development is beneficial but that niches tended to persist. We also saw that clear visions destroy constructive ambiguity. In addition clear visions seemed to hamper the development of niches, as only a lack of a clear vision on what ‘Japan’ constituted allowed some innovation actors to build a niche for their own conceptualization of ‘Japan’. Also a clearer vision, such as was the case for the robust spoor project, initiated the involvement of outside actors in the network that forced the innovation to become incremental over time. Our case study showed that systemic change was the result of multiple processes of which the non-formation of networks and the lack of a clear vision were a key part.

Secondly, the MLP needs to incorporate the notion that an innovation can change also as an idea and not solely as a concrete artifact. The involved stakeholders never tested the Japanese principles in their entirety. Most changes to the innovation happened when it was just an idea, living in the heads of involved managers, designers and other stakeholders. The idea changed from a holistic set of changes in 1997 to a more operationalized smaller set of changes in 2009 without its evolution being materialized in concrete experiments and artifacts over this period of time. However, the make up of the innovation, even with it being just an idea, had impacts on the actor networks surrounding it. Hence, theoretical frameworks need to focus on the cognitive aspects of designing an innovation as well. The MLP assumes innovation progress to occur via experiments and tests, somewhat leaving out the cognitive factor and the creative imagination of (networks of) innovation designers. This study showed however that experiments and tests are not the sole proper focal points for an analysis of the dynamics of an innovation process.

Thirdly, and closely related to the aforementioned implication, the importance of the evolution of an idea next to the evolution of the artifact points to the need to incorporate 'non-implementation' into theoretical models on implementation. We saw that the time an innovation idea was not being implemented was just as relevant for the dynamics of the implementation process as the time of actual implementation. We also saw that the innovation did not uni-directionally grow in its systemic nature, but rather lost some of its systemicity midway. Most models like the MLP deal with how an innovation gets implemented. These models start their analysis from the point where there is already an innovation to implement and see the origins of the innovation as exogenous.

Finally, our findings implicate that the MLP should incorporate the possibility and functionality of divergence of an idea. The MLP portrays an innovation process as one where multiple innovations co-align to create synergies. We saw however that from an initial set of coherent changes, the innovation diverged into a set of loosely related measures put into separate projects. This anatomy of the process, resembling the spray of a shotgun, had its functions as it allowed multiple seeds of change to find a window-of-opportunity in the regime. Also, the loosely coupled nature of the idea allowed regime players to cherry-pick elements of it and when these elements proved to be a failure, such as 'rondje om de kerk' (the assigning of train personnel to fixed routes), these failures did not impact the survivability of the innovation itself. These notions are interesting to introduce into the MLP.

Concluding on the implications of our study for the MLP we see that the MLP would benefit from incorporating into its framework two main phenomena. Firstly, the role of technological change as an endogenous factor and secondly the inherent multi-directionality of innovation processes. Regarding the latter, this thesis showed that innovations do not grow uni-directionally from a niche to regime. However, most of the concepts and assumed dynamics in the MLP embrace in them the notion of uni-directionality of implementation processes.

Innovations either die out or get implemented. We saw however that sometimes an innovation temporarily dies out but, whilst in a different form, gets implemented eventually. We depict this distinction graphically in Figure 10.1.

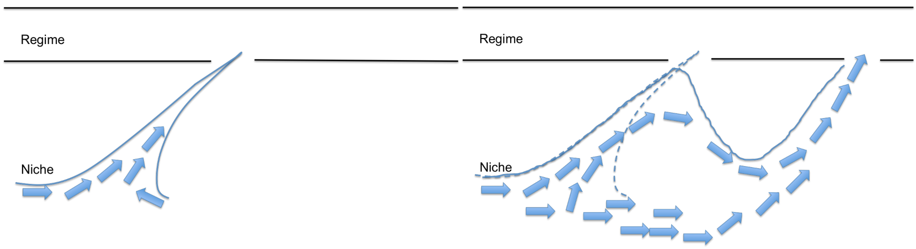


Fig 10.1 MLP's unidirectional innovation process vs. multidirectional process uncovered in this thesis

Whereas MLP looked at the structural properties of an innovation process over time, TIS looks at the innovation project itself and takes on a functional perspective. TIS-literature posits that an innovation system (a technology, actors and institutions) needs to create 7 relevant functions for an innovation to prosper. Regarding this stream of theoretical research on innovation we believe that our findings have some interesting implications.

Firstly, many of our mechanisms can be related in some way or another to the 7 functions. Interlocking for instance, is closely related to market creation as it both involves the searching for markets that can benefit from using the innovation. Whether it is a concurrent innovation process or an open market of consumers is, we believe, a matter of scale and not scope. In that sense our four mechanisms do not negate or corroborate the functions from the TIS literature. We simply took on a different perspective, used a grounded theory approach and applied the creative imagination of the researcher to arrive at four mechanisms.

However, our findings do implicate that solely looking at functions of an innovation system does not suffice. We used the PSI framework to better understand the structural properties of an innovation system (in this thesis the innovation system was the project) and we saw that structure and function are not related one-to-one. The way the structure is connected to functions is complex, meaning that if one changes the structure to create a function, one might also trigger changes in other structural elements and functions that counteract the initial benefits. In addition, functions do not reinforce each other directly, but impact each other through changes in the underlying structure. The difference in unit of analysis in TIS-research and in this thesis is depicted in Figure 10.2

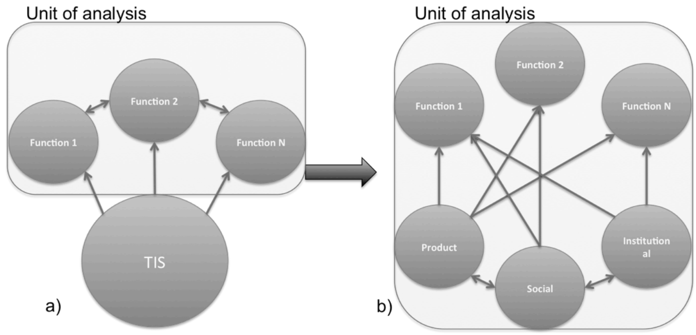


Fig. 10.2 the relation between a TIS and its functions without analyzing the structure of a TIS (a) and with the use of the PSI-framework (b)

TIS-literature conceptualizes an innovation process as the continuous reinforcing of functions. The 7 functions tend to support each other and it is this mechanism, the framework assumes, that explains the progress of an innovation process. Our findings show that functions can hinder each other. Experimenting and the creation of knowledge, one of the functions of TIS, alleviate shields and increase inclusion and subsequently the loss of niches tends to force innovations to become more incremental. More incremental innovations can become less interesting for markets because the expected benefits become fewer. This shows that market formation and knowledge creation might be functions that counteract each other. This mechanism, we believe, is due to the complex relation these functions have with the structural properties of the innovation system. These structural properties have so far not been addressed by the TIS literature. Hence, TIS would benefit from better conceptualizing and analyzing the structure of a TIS, not solely the functions. We propose the PSI-framework, but other frameworks might do as well.

Regarding methodology of both TIS and MLP, our findings implicate that studying innovation processes needs the use of multiple theoretical frameworks to do justice to the multidimensional and multi-layered complexity of these processes. For instance, innovation processes are best explained by looking both at structure and agency simultaneously. However, few theoretical frameworks allow researchers to do this. Combining TIS, MLP and NK we were better able to grasp all relevant aspects of the systemic innovation process under study. Also, academic work on innovation processes should make a better distinction between analytical frameworks and theoretical frameworks. We posit that when describing innovation processes, a less normative analytical framework prevents researchers from seeing solely what their framework mandates. We used the PSI framework as an analytical framework, which was relatively theory-free. By doing so, we found patterns that mirrored the pattern we would have expected from innovation literature. In the subsequent in-depth study of one case we did our initial analysis with a grounded theory approach, not forcing any theoretical framework on our empirical findings.

The problem with loyalty to one theoretical framework and the use of it to analyze a case study, rather than using an analytical framework, is that it becomes self-fulfilling. The framework corroborates empirical findings and vice versa. More eclecticism, by using multiple theoretical frameworks and more neutrality, by using a more descriptive analytical framework, is needed in innovation science literature.

10.6.2 Implications for managing innovation processes

We saw that the systemic innovation process was multi-layered, multi-dimensional, multi-directional and chaotic and only partially controlled by individual actors. These properties, although in some ways problematic, also helped in allowing the innovation process to progress. Our study therefore implicates that the use of projects to implement a systemic innovation can actually be counterproductive. The implementation of the Japanese principles occurred not despite serendipity but because of it. Projects would however delineate what the innovation should be and when and with whom it should be implemented and hence force one specific design of an innovation into the regime at a specific time. This consequence of using a project works against the functionality of serendipity. Serendipity allowed an innovation to connect with both a problem owner and use a window of opportunity, a connection that could not have been designed beforehand. Exemplary for this is the difference between the official Japan project, Robuust Spoor, and the DSSU project. Whereas the official project ended in only incremental changes, the DSSU project proofed to be a systemic departure from the status quo, a departure that was not designed as such beforehand.

A second implication is that innovation processes cannot be optimized. Each mechanism had its constructive and disruptive effects. Consequently, all measures taken to improve an innovation process can just as well impair an innovation process. Niches work, but they also do not work; experimentation works but it also does not work; network building works but is also does not work, etc. There are two reasons that explain this finding. Firstly, we have analyzed a transition that is much more an internal affair than a transition of a system being invaded by a competitive and parallel grown system. In that sense, this creates the effect that what is originally a political struggle between two competing systems (niche system and regime system) now becomes a choice, and hence a dilemma, for the entire system (both niche and regime). Still, we feel that the inherent dilemmas of transitions should be more thoroughly explored in current research on transitions, even if our analysis was about a transition that occurred in a more internal manner than cases usually used in the literature. Specifically we would like to point to the fact that experimentation and learning could destroy constructive ambiguity. Sometimes for a niche innovation to find the right window-of-opportunity it needs to wait for a while and mistimed experimentation, because it can cause controversy, might hinder the innovation in waiting and persisting, and using this opportunity.

Additionally, the value of concepts like joint fact-finding now becomes less self-explanatory. Increased cooperation can alleviate valuable shields and increased knowledge can alleviate valuable ambiguity. The introduction of these collaborative styles of research and development into innovation practice, something that we see increasingly occurring, should be done with careful consideration for the potential negative side effects.

10.6.3 Implications for gaming simulation literature and practice

The product development and innovation literature mainly approaches experiments, such as gaming simulation, from the analytical sciences. In addition, stakeholders in the railway sector perceived the method of gaming similarly. As we showed in Chapter 3, implicit models of sector incumbents, especially those involved in innovation, mainly resembled linear models of innovation and the analytical-scientific use of experiments. Mistakenly, stakeholders involved in employing games demand the method to deliver validated claims about the acceptance or rejection of hypotheses. In such tradition, game designers would focus highly on ecological validity and experimental design. Chapter 3 already pointed to the shortcomings of gaming simulation if perceived from this perspective.

Our further study into the use and disuse of gaming then implicated that games should be designed differently. Games are not experiments and should not be designed as such. Rather than focusing on increasing validity and reliability, game designers should design the game in such a way that they influence the four mechanisms that drive an innovation process. As we posited, validity issues can however be tackled in debriefing, thereby increasing the credibility of the outcomes to those stakeholders that need to act upon the results. However, validity measures are then solely put in place to functions as catalyst to the working of the 'active substance' of gaming: that which makes it usable to its context-of-use. In addition, whereas the process leading up to the gaming experiment and the debriefing afterwards are irrelevant if one wants to find valid causal claims, our conception of the role of gaming simulation from a design-scientific perspective forces game designers to consider the game design process itself as well. In addition, we have shown that the debriefing is the most important phase of the gaming session. This study therefore provides two new additions to the gaming simulation literature. It adds new design parameters, choices a game designer can make, and it adds new functions. The design parameters entail not only game design choices but also choices regarding the design process of the game and the design of the debriefing. The new functions entail games' impact on volatility through working on four mechanisms.

Regarding the use of games our findings implicate that gaming simulation is not a panacea nor is it a risk-free measure to take. We saw that games can influence mechanisms and that these mechanisms can be constructive or disruptive. This means that, depending on context and the actual design of the game and the debriefing, a game can actually hamper an innovation process. For instance, it might alleviate needed shielding and create controversy too soon in the innovation process. These implications arise because we have built the first step towards a better understanding of the relevant mechanisms at work in the context in which gaming simulation for innovations is employed. We saw that games can create the so-called 5C's (Duke and Geurts, 2004) but that the process not always demands these 5C's.

The scientific literature on gaming has focused mainly on the questions: why do games work? And what does this 'working' mean? The mentioned 5C-framework of Duke is the most prominent framework dealing with the active substance of gaming: games support communication, consensus, commitment to action, creativity, and the dealing with complexity. We feel that if one solely looks at games in isolation that is exactly what they do. This still holds if one looks at games for research and design, a specific application that was not the core use of gaming in Duke's work. However, this thesis additionally looked at the specific context in which gaming was applied and showed that not always the process was helped by for instance fostering creativity. Moreover, in some instances the process was severely disrupted by phenomena like too much creativity, too much consensus or too much communication. Then, taking into account that games always allow for the 5C's, albeit to different extents, games can actually hurt.

Its harmful nature, depending on the exact timing and nature of its employment, is a crucial notion for the gaming literature and has significant consequences for the design of games as well as its embedding in organizational processes. To better understand this relation between context and game design, one should therefore be a gaming professional as well as an innovation scholar. This is however problematic since the fields of gaming and the field of innovation management are quite separated. There is need for a coalescence of innovation and gaming since the working of a game cannot be seen as exogenous to the innovation process. Dynamics in the context influence the way games can be designed and can be executed, which subsequently influence dynamics in the context. Hence, game designers need to focus not only on the design of the game but also on the peculiarities of the context in which the gaming simulation session is conducted. This notion is not new (See Klabber's 2009 DIS and DIL distinction for policy-making processes) but this thesis provides more content to the actual DIS and DIL regarding innovation processes and the role of games as well as shows the notions' relevance for gaming for innovation processes.

To reiterate, the most critical design parameter of a game is the design of the debriefing afterwards. This implicates that game designers need to give more attention to the careful construction of a debriefing framework and for the facilitation of the debriefing. We provided such a framework, which needs to be refined or changed in the future as experience with debriefing games for innovation increases. In general however, the fact that it is the debriefing were the active substance of gaming materializes, demands from the game designer different qualities than the pure modeling skills used for experimental simulation games. What these skills might be was not the topic of this thesis, but it is certainly an interesting avenue for future research.

10.7 Limitations

The main conclusion resulting from this thesis is of course tentative. We have used a highly inductive and qualitative approach to our research questions since we wanted to fully grasp the complexity of the topic as well as study aspects of it that have yet to be fully conceptualized in existing theoretical frameworks. Hence, this thesis has resulted in a set of propositions about the relevant mechanisms in innovation processes and the value of gaming simulation rather than a set of tested hypotheses.

In Chapter 6 we have assessed the differences in patterns of dynamics in P, S and I spaces of three different case studies and asserted that the differences were caused by the specific makeup of the innovation. Although we could logically arrive at such a causal claim given the analysis of the case studies, this approach is highly susceptible to confounding variables. The limitations lie both in the very small sample size and the variety of methods used. Firstly, each of the three cases was different on many levels. Although this was also the reason why we chose these cases in the first place, to set aside systemic innovations from more normal innovations, this also leads to many other confounding variables potentially playing a role in explaining the differences. Different contextual factors such as culture and more formal institutions have differed per case and we cannot fully rule out their impact on the different patterns we have observed. Secondly, we had to apply different observation methods to uncover the patterns due to practical concerns. Whereas for the spoorzone Delft case and to a lesser extent the DSSU case we could rely on direct observations, the Network Rail case fully depended on retrospective accounts of interview respondents. Then this specific case might be more susceptible to post-hoc rationalization of respondents.

In Chapter 7 we delved deeper in one specific case to uncover the mechanisms that drive the pattern we had found in the previous chapter. Although single case studies are most suitable for more in-depth analysis, this came at the cost of generalizability. To what extent do our findings from one case apply to other similar cases? Albeit that this is a serious shortcoming of single case studies, we feel that we have overcome this limitation partly by making our causal claims less specific and more general. We have stated that four mechanisms can be found and that they play a constructive and disruptive role in innovation processes. This highly descriptive observation, rather than a more predictive one, makes it less susceptible to generalizability issues.

Probably a bigger limitation to the single case study is the grounded theory approach that we used for making sense of the data. This approach always involves some creative interpretation of the researcher since the data never fully fits one resulting theory. We have tried to make this creative interpretation as tractable and transparent as possible but the very fact of it always results in demanding from the reader a judgmental leap from data to conclusions. In addition, our analysis of the data and the path to the resulting conclusions might be sensitive to initial conditions. Had we looked differently to the data from the beginning, through different lenses or focusing on different aspects of the data, would we then have arrived at different mechanisms later on?

In Chapter 8 we addressed the role of gaming in impacting the mechanisms of Chapter 7 and the patterns in Chapter 6. Whereas the previous chapters relied heavily on a very limited set of cases, this chapter encompassed an analysis of around ten games. For quantitative analysis still a too small sample, but for finding the dominant parameters of the causal model linking games and innovation processes, this was sufficient. Still, we have to be very careful in stating our conclusions regarding gaming. 10 cases still leave open the possibility of confounding variables explaining its effect on the mechanisms rather than the game, or game design elements, itself.

Subsequently we went on designing a framework for debriefing games, since we stated that the debriefing needs far more attention than it currently gets. This framework is highly normative yet still never used and tested in its entirety. This poses a true limitation to our claims in this chapter, since we have only based our framework on lessons learned over the many games we have employed and debriefed and not on actually using the framework itself. The reason for it being that since we have both found the crucial importance of debriefing and have built the framework after our analysis of the games and subsequently weren't able to design any new games.

Overarching to the limitations per chapter is the buildup of this thesis: sequential rather than parallel. Each chapter provided the input for the subsequent chapter. This makes the thesis prone to small changes in the validity of propositions causing the entire thesis to become faulty. If for instance, a new conceptualization of systemic innovation processes' difference from other processes results in something different from our 'fuzzy back-end'-claim, do we arrive at a set of different mechanisms? To some extent this is problematic although we can posit that each chapter became more specific and increased in level in detail thereby causing the early chapters to be less prone to falsification. In addition, other conceptions of relevant macro-level patterns will be more of an addition to the analysis of the role of games than a refutation of our claims.

The abovementioned limitations are serious and provide many opportunities for further research in this area to improve, or to build on the work that has been done in this thesis. Still we feel that the true contribution of this thesis is not severely impacted by the limitations. What we have tried to do is show existing theory on innovation and games to what extent the messiness of reality fits the many theoretical models abound in the scientific literature. Our contribution lies in stating that systemic innovation processes and the use of gaming is not that easy as many models implicitly or explicitly assume. This contribution is more modest than providing exact causal claims about the value of gaming and therefore also less impacted by the limitations of this research.

10.8 Future research

Most obviously future research could focus on replicating the mechanisms found in other cases where a sociotechnical system is trying to change its evolutionary path. This thesis has used the railway sector but similar processes are to be expected in sectors like energy, water, waste sectors. In addition, transitions of another kind could be the focus of study, looking at the extent to which the mechanisms also play a role there. These transitions might be over longer periods of time, more open-ended, or involve different type of innovations.

Also, more rigorous methodological studies could be used to not replicate the findings but rather test them. Then, it would involve more fully operationalizing the concepts that this study provided and measuring them in a more controlled setting. Although we feel that this is cumbersome given the context-specificity of innovation processes in real life, an attempt could very well prove to be valuable in refining and validating the research outcomes presented here.

A better operationalizing of the constructs could also be helpful for the practical relevance of this thesis. We have provided a set of mechanisms and an overview of the ways gaming simulation can influence these. However, for the practitioner there is still some fuzziness on how to determine whether a mechanism is present and a specific point in time and how to design a game such that it will impact this mechanism. This first and foremost has to do with the fact that most pivotal points in the timeline of the DSSU process were used to construct an overarching theory on how these processes move forward, linking many pivotal points to each other. Hence we forwent on more deeply analyzing the true activities, strategies and incentives at play at these specific pivotal points. Such a study might be interesting since it could shed more light on the specific use of gaming and might increase the likelihood that the gaming practitioner is able to diagnose the mechanisms at play during a specific situation. These mechanisms might be the mechanisms uncovered in this thesis, but they may very well be of a different order given the more micro-level unit of analysis. An interesting perspective might be game theory as during these pivotal points many parties with different strategies congregate and try to influence the end-result of the process. In addition, by portraying gaming simulation environments as niches, we assumed that such environments were free of the usual institutional pressures that forbid all too radical change. However, a more in-depth look at what rationales, incentives, norms and values are at play during a gaming simulation sessions might show that rather than being free of 'institutions', simply other institutions apply. What these other institutions are is an interesting topic for further research.

In addition, we feel that the debriefing framework could be enriched by games applied in different context as well as tested in its entirety. Games for policy making or games for pure experimental research probably demand different approaches to the debriefing and it is interesting to see to what extent the framework presented in this thesis is still valuable. In addition, the skills needed to debrief properly are underexplored in this thesis. The debriefing framework of this thesis focuses on more on substance. The question how to facilitate a debriefing is an interesting on. In general, the topic of debriefing should be a great avenue for further research as this is a crucial yet underdeveloped part of gaming simulation for research.

Bibliography

Abbott, A. (2001). *Time matters: On theory and method*. Chicago: University of Chicago Press.

Abernathy, W.J., & Utterback, J.M. (1978). Patterns of industrial innovation. *Technology review*, 80(7), 40-47.

Abril, M., Barber, F., Ingolotti, L., Salido, M.A., Tormos, P., & Lova, A. (2008). An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, 44(5), 774-806.

Ackoff, R.L. (1971). Towards a system of systems concepts. *Management science*, 17(11), 661-671.

Ackoff, R.L. (1974). The systems revolution. *Long Range Planning*, 7(6), 2-20.

Ackoff, R.L. (1979). The future of operational research is past. *Journal of the operational research society*, 30(2), 93-104.

Alkemade, F., Frenken, K., Hekkert, M.P., & Schwoon, M. (2009). A complex systems methodology to transition management. *Journal of Evolutionary Economics*, 19(4), 527.

Altenberg, L., 1995. Genome growth and the evolution of the genotype–phenotype map. In: Banzhaf, W., Eckman, F.H. (Eds), *Evolution and Biocomputation*. Berlin: Springer-Verlag, 205–259

Anderson, N., De Dreu, C.K., & Nijstad, B.A. (2004). The routinization of innovation research: a constructively critical review of the state of the science. *Journal of organizational Behavior*, 25(2), 147-173.

Anderson Jr, E.G., & Joglekar, N.R. (2012). *The innovation butterfly: Managing emergent opportunities and risks during distributed innovation*. New York: Springer.

Ashby, W.R. (1956). *An Introduction to cybernetics*. London: Chapman & Hall.

Axelrod, R. (1997). Advancing the art of simulation in the social sciences. In Conte, R., Hegselmann, R. & P. Terna (Eds) *Simulating social phenomena*. Berlin: Springer Verlag, 21-40

Balci, O. (1998). Verification, validation, and accreditation. In Medeiros, D.J., Watson, E.F., Carson, J.S., & M.S. Manivannan (Eds). *Proceedings of the 30th conference on Winter simulation*. Piscataway: IEEE Computer Society Press. 4-41

Barreteau, O., Bousquet, F., & Attonaty, J. M. (2001). Role-playing games for opening the black box of multi-agent systems: method and lessons of its application to Senegal River Valley irrigated systems. *Journal of artificial societies and social simulation*, *4*(2), 5-18

Bekebrede, G., & Mayer, I. (2006). Build your seaport in a game and learn about complex systems. *Journal of design research*, 5(2), 273-298.

Bergek, A., Jacobsson, S., Carlsson, B., Lindmark, S. & Rickne, A. (2008). Analyzing the functional dynamics of technological innovation systems: A scheme of analysis. *Research policy*, 37(3), 407-429.

Berkhout, F., Smith, A. & Stirling, A. (2004). Socio-technological regimes and transition contexts. *System innovation and the transition to sustainability: theory, evidence and policy*, 44(106), 48-75.

Beunza, D. & Stark, D. (2004). Tools of the trade: the socio-technology of arbitrage in a Wall Street trading room. *Industrial and corporate change*, 13(2), 369-400.

Bizzi, L., & Langley, A. (2012). Studying processes in and around networks. *Industrial Marketing Management*, 41(2), 224-234.

Bontekoning, Y., & Priemus, H. (2004). Breakthrough innovations in intermodal freight transport. *Transportation Planning and Technology*, 27(5), 335-345.

Bouwmans, I., Weijnen, M.P.C., & Gheorghe, A.V. (2006). Infrastructures at risk. In Georghe. A.V., Masera, M., Weijnen, M.P.C., De Vries, L. (Eds) *Critical Infrastructures at Risk*. Amsterdam: Springer Netherlands. 19-36

Bratley, P., Fox, B.L., & Schrage, L.E. (2011). *A guide to simulation*. Heidelberg: Springer Science & Business Media.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta psychologica*, 81(3), 211-241.

Brown, S.L., & Eisenhardt, K.M. (1995). Product development: Past research, present findings, and future directions. *Academy of management review*, 20(2), 343-378.

De Bruijn, J.A. & Herder, P.M. (2009). System and Actor Perspectives on Sociotechnical Systems. *Systems, Man and Cybernetics, Part A: Systems and Humans*, 39, (5) 981-992

Brusoni, S., & Prencipe, A. (2001). Unpacking the black box of modularity: technologies, products and organizations. *Industrial and Corporate Change*, 10(1), 179-205.

Bucciarelli, L.L. (1984). Reflective practice in engineering design. *Design studies*, 5(3), 185-190.

Burns, T.E., & Stalker, G.M. (1961). *The management of innovation*. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship. Available at SSRN: https://ssrn.com/abstract=1496187

Bush, V. (1945). As we may think. *The atlantic monthly*, 176(1), 101-108.

Campbell, D.T., & Stanley, J.C. (2015). *Experimental and quasi-experimental designs for research*. Memphis: Ravenio Books.

Caniels, M.C., & Romijn, H.A. (2008). Strategic niche management: towards a policy tool for sustainable development. *Technology Analysis and Strategic Management*, 20(2), 245-266.

Carlsson, B., & Jacobsson, S. (1997). In search of useful public policies—key lessons and issues for policy makers. In Carlsson, B. (Ed) *Technological systems and industrial dynamics.* New York: Springer. 299-315

Carlsson, B., & Stankiewicz, R. (1991). On the nature, function and composition of technological systems. *Journal of evolutionary economics*, 1(2), 93-118.

Cheng, Y.T., & Van de Ven, A.H. (1996). Learning the innovation journey: Order out of chaos? *Organization science*, 7(6), 593-614.

Chesbrough, H. (2003). The logic of open innovation: managing intellectual property. *California Management Review*, 45(3), 33-58.

Christensen, C. (1997). *The innovator's dilemma: when new technologies cause great firms to fail*. Cambridge: Harvard Business Review Press.

Fujimoto, T., & Clark, K. B. (1991). *Product development performance: Strategy, organization, and management in the world auto industry*. Boston: Harvard business school press

Cohen, M.D., March, J.G., & Olsen, J.P. (1972). A garbage can model of organizational choice. *Administrative science quarterly*, 1-25.

Cohen, W.M., & Levinthal, D.A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, 128-152.

Cooper, R.G. (1990). Stage-gate systems: a new tool for managing new products. *Business horizons*, 33(3), 44-54.

Cooper, R.G. (1994). Perspective third-generation new product processes. *Journal of product innovation management*, 11(1), 3-14.

Cooper, R.G. (2008). Perspective: The Stage-Gate® idea-to-launch process—Update, what's new, and NexGen systems. *Journal of product innovation management*, 25(3), 213-232.

Creswell, J. (2003). *Research design: qualitative, quantitative and mixed methods*. Thousand Oaks: Sage

Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & gaming*, 41(6), 898-920.

Crozier, M. (2009). *The bureaucratic phenomenon*. New Brunswick: Transaction Publishers.

Cyert, R.M., & March, J.G. (1963). *A behavioral theory of the firm*. Cambridge: Blackwell

Dalton, M. (1959) *Men who manage*. New York: Wiley

Damanpour, F. (1991). Organizational innovation: A meta-analysis of effects of determinants and moderators. *Academy of management journal*, 34(3), 555-590.

Davies, A., & Hobday, M. (2005). *The business of projects: managing innovation in complex products and systems*. Cambridge: Cambridge University Press.

De Bruijn, J.A. & Herder, P.M. (2009). System and Actor Perspectives on Sociotechnical Systems. *Systems, Man and Cybernetics, Part A: Systems and Humans*, 39, (5) 981-992

De Caluwé, L. (1997). *Veranderen Moet je Leren. Een Evaluatiestudie naar de Opzet en Effecten van een Grootschalige Cultuurinterventie met Behulp van Spelsimulatie*. (in Dutch), Den Haag: Delwel

De Neufville, R., & Scholtes, S. (2011). *Flexibility in engineering design*. MIT Press.

De Weck, O. ., Roos, D., & Magee, C.L. (2011). *Engineering systems: Meeting human needs in a complex technological world*. Cambridge: MIT Press.

Decker, S., Fey, M., Sideras, S., Caballero, S., Boese, T., Franklin, A.E. & Borum, J.C. (2013). Standards of best practice: Simulation standard VI: The debriefing process. *Clinical Simulation in Nursing*, 9(6), S26-S29.

Dennehy, R.F., Sims, R.R., & Collins, H.E. (1998). Debriefing experiential learning exercises: A theoretical and practical guide for success. *Journal of Management Education*, 22(1), 9-25.

Department for Transport (2011). *Realising the potential of GB rail: an independent report of a study chaired by Sir Roy McNulty and commissioned by the Secretary of State for Transport.* London: Department for Transport

Diehl, E., & Sterman, J.D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 62(2), 198-215.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*, 11(3), 147-162.

Drack, M. (2009). Ludwig von Bertalanffy's early system approach. *Systems Research and Behavioral Science*, 26(5), 563-572.

Dubois, A., & Gadde, L.E. (2002). Systematic combining: an abductive approach to case research. *Journal of business research*, 55(7), 553-560.

Duin, H., Baalsrud Hauge, J., Thoben, K.D. & Bierwolf, R.E. (2009). A Concept for Serious Gaming to Support Disruptive Idea Generation. In: Thoben, K.D., Pawar, K.S., Katzy, B.R. & Bierwolf, R.E. (Eds) *Proceedings of the 15th International Conference on Concurrent Engineering*. Nottingham: Centre for Concurrent Enterprise. 22-24

Duke, R.D. (2011). Origin and evolution of policy simulation: A personal journey. *Simulation & Gaming*, 42(3), 342 - 358

Duke R.D., & Geurts, J.L.A. (2004). *Policy Games for Strategic Management*. Amsterdam: Dutch University Press

Edquist, C. (1997). *Systems of innovation: technologies, institutions, and organizations*. London: Routlegde

Egyedi, T.M., & Mehos, D.C. (2012). *Inverse Infrastructures: Disrupting networks from below*. Cheltenham: Edward Elgar Publishing.

Ehrman J.R. & Stinson, B.L. (1999). Joint fact-finding and the use of technical experts, In Susskind, L., McKearnan, S., Thomas-Larmer, J. (Eds), *The Consensus Building Handbook*, Thousand Oaks: Sage. 375–399

Eisenhardt, K.M. (1989). Building theories from case study research. *Academy of management review*, 14(4), 532-550.

Emery, F.E. (1959) *Characteristics of sociotechnical systems*. London: Tavistock

Englemore, R.S., & Tenenbaum, J.M. (1990). *The Engineers Associate: ISAT Summer Study Report*. Unpublished Report, Palo Alto: Stanford University.

Erat, S., & Kavadias, S. (2008). Sequential testing of product designs: Implications for learning. *Management Science*, 54(5), 956-968.

Ethiraj, S.K., Levinthal, D., & Roy, R.R. (2008). The dual role of modularity: innovation and imitation. *Management Science*, 54(5), 939-955.

Ethiraj, S.K., & Levinthal, D. (2004). Modularity and innovation in complex systems. *Management science*, 50(2), 159-173.

Etzioni, A. (1967). Mixed-Scanning: a third approach to decision-making. *Public Administration Review*. 27(5) 385-392

Fagerberg, J. (2005). *The Oxford handbook of innovation*. Oxford: Oxford university press.

Fanning, R.M., & Gaba, D.M. (2007). The role of debriefing in simulation-based learning. *Simulation in healthcare*, 2(2), 115-125.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *Journal of Problem Solving*, 4(1), 19-42

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. *Research policy*, 30(7), 1019-1039.

Forrester, J.W. 1958. "Industrial Dynamics: a Major Breakthrough for Decision Makers." *Harvard Business Review*, Vol. 36, No. 4, 37-66

Frantzeskaki, N., & Loorbach, D. (2010). Towards governing infrasystem transitions: reinforcing lock-in or facilitating change? *Technological Forecasting and Social Change*, 77(8), 1292-1301.

Freeman, M.J. & Aldcroft D.H. (1985). *The Atlas of British Railway History*, Beckenham: Croom Helm

Frenken, K. (2000). A complexity approach to innovation networks. The case of the aircraft industry (1909–1997). *Research Policy*, 29(2), 257-272.

Frenken, K. (2006). *Innovation, evolution and complexity theory*. Cheltenham: Edward Elgar Publishing.

Frenken, K., & Mendritzki, S. (2012). Optimal modularity: a demonstration of the evolutionary advantage of modular architectures. *Journal of Evolutionary Economics*, 22(5), 935-956.

Frigg, R., & Reiss, J. (2009). The philosophy of simulation: hot new issues or same old stew? *Synthese*, 169(3), 593-613.

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In Sternberg R.J. & Frensch, P.A. (Eds) *Complex problem solving: Principles and mechanisms*. New York: Psychology Press 185-222.

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning*, 7(1), 69-89.

Funke, J. (2003). *Problemlosendes Denken* [Problem solving]. Stuttgart: Kohlhammer.

Galbraith, J.R. (1977). *Organization design*. Boston: Addison Wesley Publishing Company.

Garud, R., & Van de Ven, A.H. (2002). Strategic change processes. In Pettigrew, A., Thomas, H. & Whittington, R. (Eds) *Handbook of strategy and management*, Thousand Oaks: Sage. 206-231.

Gavetti, G., Levinthal, D.A., & Rivkin, J.W. (2005). Strategy making in novel and complex worlds: The power of analogy. *Strategic Management Journal*, 26(8), 691-712.

Gavetti, G., & Levinthal, D.A. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative science quarterly*, 45(1), 113-137.

Geels, F.W. 2002. Technological Transitions as Evolutionary Reconfiguration Processes: a Multi-level Perspective and a Case-study, *Research Policy* 31(8): 1257-1274.

Geels, F.W. (2004). From sectoral systems of innovation to socio-technical systems: Insights about dynamics and change from sociology and institutional theory. *Research policy*, 33(6), 897-920.

Geels, F.W. (2005). Processes and patterns in transitions and system innovations: refining the co-evolutionary multi-level perspective. *Technological forecasting and social change*, 72(6), 681-696.

Geels, F.W. (2005). The dynamics of transitions in socio-technical systems: a multi-level analysis of the transition pathway from horse-drawn carriages to automobiles (1860–1930). *Technology analysis & strategic management*, 17(4), 445-476.

Geels, F.W. (2010). Ontologies, socio-technical transitions (to sustainability), and the multi-level perspective. *Research policy*, 39(4), 495-510.

Geels, F.W. (2011). The multi-level perspective on sustainability transitions: Responses to seven criticisms. *Environmental innovation and societal transitions*, 1(1), 24-40.

Geels, F.W. (2012). A socio-technical analysis of low-carbon transitions: introducing the multi-level perspective into transport studies. *Journal of Transport Geography*, 24, 471-482.

Geels, F.W., & Schot, J. (2007). Typology of sociotechnical transition pathways. *Research policy*, 36(3), 399-417.

Genus, A. & Coles, A.M. (2008). Rethinking the multi-level perspective of technological transitions. *Research Policy*, 37(9), 1436-1445.

George, A.L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge: MIT Press.
Geurts, J.L., & Joldersma, C. (2001). Methodology for participatory policy analysis. *European Journal of Operational Research*, 128(2), 300-310.
Geyer, A., & Davies, A. (2000). Managing project–system interfaces: case studies of railway projects in restructured UK and German markets. *Research Policy*, 29(7), 991-1013.
Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis*. Berkeley: University of California Press.
Godin, B. (2006). The linear model of innovation: The historical construction of an analytical framework. *Science, Technology, & Human Values*, 31(6), 639-667.
Gökalp, I. (1992). On the analysis of large technical systems. *Science, Technology, & Human Values*, 17(1), 57-78.
Gopalakrishnan, S., & Damanpour, F. (1997). A review of innovation research in economics, sociology and technology management. *Omega*, 25(1), 15-28.
Goverde, R.M.P. (2005) *Punctuality of railway operations and timetable stability analysis*. Delft: TU Delft repository
Greif, S., & Kurtz, H.J. (1996). *Handbuch Selbstorganisiertes Lernen*. Göttingen: Verlag für Angewandte Psychologie.
Greitzer, F.L., Podmore, R., Robinson, M., & Ey, P. (2010). Naturalistic decision making for power system operators. *International Journal of Human–Computer Interaction*, 26(2-3), 278-291.
Griffin, L.J. (1993). Narrative, event-structure analysis, and causal interpretation in historical sociology. *American journal of Sociology*, 98(5), 1094-1133.
Grin, J., Rotmans, J., & Schot, J. (2011). On patterns and agency in transition dynamics: Some key insights from the KSI programme. *Environmental Innovation and Societal Transitions*, 1(1), 76-81.
Hage, J. (1980). *Theories of organizations: Form, process, and transformation*. New York: John Wiley & Sons.
Hage, J., & Aiken, M. (1969). Routine technology, social structure, and organization goals. *Administrative science quarterly*. 14(3), 366-376.
Hambrick, D.C., & Crozier, L.M. (1985). Stumblers and stars in the management of rapid growth. *Journal of Business Venturing*, 1(1), 31-45.
Hardy, C. (1994). *Managing strategic action: mobilizing change: concepts, readings, and cases*. Thousand Oaks: Sage
Hardy, C., Lawrence, T.B., & Grant, D. (2005). Discourse and collaboration: The role of conversations and collective identity. *Academy of management review*, 30(1), 58-77.
Harrison, G.W., & List, J.A. (2004). Field experiments. *Journal of Economic literature*, 42(4), 1009-1055.
Harteveld, C., Guimarães, R., Mayer, I.S., & Bidarra, R. (2009). Balancing play, meaning and reality: The design philosophy of LEVEE PATROLLER. *Simulation & Gaming*. 41(3), 316-340
Hatch, D. (2000). Cooperative Benchmarking between the Dutch and Japanese Railways. *Japan Railway & Transport Review*, 7, 18-25.
Hedström, P., & Bearman, P. (2009). *The Oxford handbook of analytical sociology*. Oxford: Oxford University Press.
Heise, D.R. (1989). Modeling event structures. *Journal of Mathematical Sociology*, 14(2-3), 139-169.
Hekkert, M.P., Suurs, R.A., Negro, S.O., Kuhlmann, S., & Smits, R.E. (2007). Functions of innovation systems: A new approach for analysing technological change. *Technological forecasting and social change*, 74(4), 413-432.
Hekkert, M.P., & Negro, S.O. (2009). Functions of innovation systems as a framework to understand sustainable technological change: Empirical evidence for earlier claims. *Technological forecasting and social change*, 76(4), 584-594.
Henderson, R.M., & Clark, K.B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.
Herder, P.M., Bouwmans, I., Dijkema, G.P., Stikkelman, R.M., & Weijnen, M.P.G. (2008). Designing infrastructures using a complex systems perspective. *Journal of design research*, 7(1), 17-34.
Hermans, F., van Apeldoorn, D., Stuiver, M., & Kok, K. (2013). Niches and networks: Explaining network evolution through niche formation processes. *Research Policy*, 42(3), 613-623.
Hofstede, G.J., De Caluwé, L., & Peters, V. (2010). Why simulation games work - in search of the active substance: A synthesis. *Simulation & Gaming,* 41(6), 824-843
Hogarth, R.M., & Reder, M.W. (1987). *Rational choice: The contrast between economics and psychology*. Chicago: University of Chicago Press.
Hoholm, T. (2011). *The contrary forces of innovation: An ethnography of innovation in the food industry*. New York: Springer.
Hoholm, T., & Araujo, L. (2011). Studying innovation processes in real-time: The promises and challenges of ethnography. *Industrial Marketing Management*, 40(6), 933-939.
Holland, J.H. (1992). Complex adaptive systems. *Daedalus*, 121(1), 17-30.
Holland, J.H. (2006). Studying complex adaptive systems. *Journal of systems science and complexity*, 19(1), 1-8.
Holland, J.H. & Miller, J.H. (1991). Artificial Adaptive Agents in Economic Theory, *American Economic Review*, 81(2), 365-370.

Hoogma, R., Kemp, R., Schot, J. & Truffer, B. (2002). *Experimenting for sustainable transport: the approach of strategic niche management*. London: Taylor & Francis.

Hounshell, D.A., & Smith, J.K. (1988). *Science and Corporate Strategy: Du Pont R and D, 1902-1980*. Cambridge: Cambridge University Press.

Howard-Grenville, J.A., & Carlile, P.R. (2006). The incompatibility of knowledge regimes: Consequences of the material world for cross-domain work. *European Journal of Information Systems*, 15(5), 473-485.

Hughes, T.P. (1983). *Networks of Power, Electrification in Western Society, 1880–1930*. Baltimore: Johns Hopkins University Press

Hughes, T.P. (1987). The evolution of large technological systems. In Bijker, W.E., Hughes, T.P., Pinch, T. (Eds) *The social construction of technological systems: New directions in the sociology and history of technology*, Cambridge: MIT Press, 51-82.

Joerges, B. (1988). Technology in everyday life: conceptual queries. *Journal for the Theory of Social Behaviour*, 18(2), 219-237.

Joldersma, C. & Geurts, J.L.A. (1998). Simulation/Gaming for Policy Development and Organizational Change. *Simulation & Gaming*, 29(4), 391-399

Jonsson, D.K. (2005). The nature of infrasystem services. *Journal of infrastructure systems*, 11(1), 2-8.

Kanter, R.M. (1988). Three tiers for innovation research. *Communication Research,* 15(5), 509-523.

Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of management journal*, 45(6), 1183-1194.

Kauffman, S.A. (1993). *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA.

Kauffman, S.A., Lobo, J., & Macready, W.G. (2000). Optimal search on a technology landscape. *Journal of Economic Behavior & Organization*, 43(2), 141-166.

Kauffman, S.A., & Macready, W. (1995). Technological evolution and adaptive organizations: Ideas from biology may find applications in economics. *Complexity*, 1(2), 26-43.

Kemp, R., Schot, J., & Hoogma, R. (1998). Regime shifts to sustainability through processes of niche formation: the approach of strategic niche management. *Technology analysis & strategic management*, 10(2), 175-198.

Kemp, R., Rip, A., Schot, J.W., 2001. Constructing transition paths through the management of niches. In: Garud, R., Karnoe, P. (Eds.), *Path Dependence and Creation*. Mahwah: Lawrence Erlbaum. 269–299.

Kemp, R., & Rotmans, J. (2005). The management of the co-evolution of technical, environmental and social systems. In Weber, M., Hemmelskamp, J. (Eds) *Towards environmental innovation systems*. Heidelberg: Springer. 33-55

Klabbers, J.H. (1993). Gaming Concerning Management of Change. In: Percival, F., Lodge, S., & Saunders, D. (Eds.) *Developing Transferable Skills in Education and Training. The Simulation & Gaming: An International Journal Yearbook*. London: Kogan Page. 130-130

Klabbers, J.H. (2003). Gaming and simulation: Principles of a science of design. *Simulation & Gaming*, 34(4), 569-591.

Klabbers, J.H. (2006). A framework for artifact assessment and theory testing. *Simulation & Gaming*, 37(2), 155-173.

Klabbers, J.H. (2009) *The Magic Circle: principles of gaming and simulation. [3rd en rev. ed.]* Rotterdam: Sense Publishers

Klabbers, J. H. (2018). On the architecture of game science. *Simulation & Gaming*, 49(3) 207-245

Klein, G.A. (1989). Recognition-primed decisions. In W. Rouse (Ed.) *Advances in man-machine systems research*. Greenwich: JAI Press. 47–92

Woolthuis, R.K., Lankhuizen, M., & Gilsing, V. (2005). A system failure framework for innovation policy design. *Technovation*, 25(6), 609-619.

Kline, S.J. (1985). Innovation is not a linear process. *Research management*, 28(4), 36-45.

Kline, J.K., & Rosenberg, N. (1986). An Overview of Innovation. In Landau, R. & Rosenburg, N. (Eds) *The Positive Sum Strategy: Harnessing Technology for Economic Growth*. Washington: National Academy Press. 275 - 306

Knudsen, T., & Levinthal, D.A. (2007). Two faces of search: Alternative generation and alternative evaluation. *Organization Science*, 18(1), 39-54.

Kolb, D.A. (2014). *Experiential learning: Experience as the source of learning and development*. Upper Saddle River: Pearson

Konrad, K., Truffer, B., & Voß, J.P. (2008). Multi-regime dynamics in the analysis of sectoral transformation potentials: evidence from German utility sectors. *Journal of Cleaner Production*, 16(11), 1190-1202.

Koppenjan, J.F.M., & Groenewegen, J. (2005). Institutional design for complex technological systems. *International Journal of Technology, Policy and Management*, 5(3), 240-257.

Koppenjan, J.F.M., & Klijn, E.H. (2004). *Managing uncertainties in networks: a network approach to problem solving and decision making*. London: Routledge

Kornish, L.J., & Ulrich, K.T. (2011). Opportunity spaces in innovation: Empirical analysis of large samples of ideas. *Management Science*, 57(1), 107-128.

Korsten, A.F.A., J.A. de Bruijn, P. de Jong and W.P.C. van Zanten (1996) Technischcomplexe projecten: een verkenning [technically complex projects: an exploration] In De Bruijn, J.A., De Jong, P., Korsten, A.F.A. & Van Zanten, W.P.C (Eds) *Grote Projecten: besluitvorming en management*. Alphen a/d Rijn: Samson. 22-105

Kriz, W.C. (2003). Creating effective learning environments and learning organizations through gaming simulation design. *Simulation & Gaming,* (34)4, 495-511.

Kriz, W.C. (2010). A systemic-constructivist approach to the facilitation and debriefing of simulations and games. *Simulation & Gaming*, 41(5), 663-680.

Kriz, W.C., & Hense, J.U. (2006). Theory-oriented evaluation for the design of and research in gaming and simulation. *Simulation & Gaming*, 37(2), 268-283.

Ksoll, M. (2004,). Integration of infrastructure and transport: An assessment from industrial economics and railway perspectives. *Paper presented at the 2nd Conference on Railroad Industry Structure, Competition and Investment*. October 8-9 2004, Evanston, IL.

Künneke, R., Groenewegen, J., & Ménard, C. (2010). Aligning modes of organization with technology: Critical transactions in the reform of infrastructures. *Journal of Economic Behavior & Organization*, 75(3), 494-505.

Langer, E.J. (1975). The illusion of control. *Journal of personality and social psychology*, 32(2), 311.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management review*, 24(4), 691-710.

Langley, A. (2007). Process thinking in strategic organization. *Strategic organization*, 5(3), 271-282.

Langley, A., Smallman, C., Tsoukas, H., & Van de Ven, A.H. (2013). Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal*, 56(1), 1-13.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge: Harvard university press.

Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale welt*, 47(4), 369-381.

Laursen, K., & Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. *Strategic management journal*, 27(2), 131-150.

Lawrence, P.R., & Lorsch, J.W. (1967). Differentiation and integration in complex organizations. *Administrative science quarterly*, 12(1), 1-47.

Leavitt, H.J. (1965). Applied organizational change in industry: Structural, technological and humanistic approaches. In J.G. March (Ed) *Handbook of organizations*, Chicago: Rand McNally, 1144-1170.

Lederman, L.C. (1992). Debriefing: Toward a systematic assessment of theory and practice. *Simulation & gaming*, 23(2), 145-160.

Lederman, L.C., & Stewart, L.P. (1987). The Marble Company: The design and implementation of a simulation board game. *Simulation & Gaming*, 18(1), 57-81.

Leonard-Barton, D. (1988). Implementation as mutual adaptation of technology and organization. *Research policy*, 17(5), 251-267.

Levine, M. (1971). Hypothesis theory and nonlearning despite ideal SR-reinforcement contingencies. *Psychological review*, 78(2), 130.

Levinthal, D.A. (1997). Adaptation on rugged landscapes. *Management science*, 43(7), 934-950.

Levinthal, D.A. (1998). The slow pace of rapid technological change: gradualism and punctuation in technological change. *Industrial and corporate change*, 7(2), 217-247.

Levinthal, D.A., & Warglien, M. (1999). Landscape design: Designing for local action in complex worlds. *Organization Science*, 10(3), 342-357.

Lindblom, C.E. (1979). Still muddling, not yet through. *Public administration review*, 39(6), 517-526.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational behavior and human decision processes*, 69(2), 149-163.

Lo, J., Van den Hoogen, J., & Meijer, S.A. (2013). Using gaming simulation experiments to test railway innovations: implications for validity. In Hill, R., Kuhl, M., Pasupathy, R., Kim, S. & Tolk, A. (Eds) *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*. Piscataway: IEEE Press. 1766-1777

Loch, C.H., Terwiesch, C., & Thomke, S. (2001). Parallel and sequential testing of design alternatives. *Management Science*, 47(5), 663-678.

Lovell, K. (2012) Rebuilding after the storm: the role of innovation in completing infrastructure system transformation. Paper presented at the Druid Academy 2012, January 19-21, Aalborg, Denmark

Lovell, K., Bouch, C., Smith, A., Nash, C., Roberts, C., Wheat, P., & Smith, R. (2011). Introducing new technology to the railway industry: system-wide incentives and impacts. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 225(2), 192-201.

Lundvall, B.Å. (1998). Why study national systems and national styles of innovation? *Technology analysis & strategic management*, 10(4), 403-422.

Malerba, F. (2002). Sectoral systems of innovation and production. *Research policy*, 31(2), 247-264.

Manzo, G. (2010). Analytical sociology and its critics. *European Journal of Sociology*, 51(01), 129-170.

March, J.G. (1991). Exploration and exploitation in organizational learning. *Organization science*, 2(1), 71-87.

March, J.G. & Simon, H.A. (1958) *Organizations*. Oxford: Wiley Organizations

Markard, J. (2011). Transformation of infrastructures: sector characteristics and implications for fundamental change. *Journal of Infrastructure Systems*, 17(3), 107-117.

Markard, J., & Truffer, B. (2006). Innovation processes in large technical systems: Market liberalization as a driver for radical change? *Research Policy*, 35(5), 609-625.

Markard, J., & Truffer, B. (2008). Actor-oriented analysis of innovation systems: exploring micro–meso level linkages in the case of stationary fuel cells. *Technology Analysis & Strategic Management*, 20(4), 443-464.

Markard, J., & Truffer, B. (2008b). Technological innovation systems and the multi-level perspective: Towards an integrated framework. *Research policy*, 37(4), 596-615.

Marquis, D.G. (1969). The anatomy of successful innovations. Innovation, 1(7), 28-37.

Mason, R.O., & Mitroff, I.I. (1981). *Challenging strategic planning assumptions: Theory, cases, and techniques*. New York: Wiley.

Mayer, I.S. (2009). The gaming of policy and the politics of gaming: A review. *Simulation & Gaming*, 40(6), 825-862.

Meelen, T., & Farla, J. (2013). Towards an integrated framework for analysing sustainable innovation policy. *Technology Analysis & Strategic Management*, 25(8), 957-970.

Meijer, S.A. (2009). *The organisation of transactions - Studying supply networks using gaming simulation*. Wagening: Wageningen Academic Publishers

Meijer, S.A. (2012). Gaming simulations for railways: Lessons learned from modeling six games for the Dutch infrastructure management. In Perpinya X. (Ed) *Infrastructure Design, Signalling and Security in Railway*. Rijeka: InTech 275-294

Meijer, S.A. (2015). The power of sponges: Comparing high-tech and low-tech gaming for innovation. *Simulation & Gaming*, 46(5), 512-535.

Meijer, S.A., Y. Reich, and E. Subrahmanian. (2014). The Future of Gaming for Design of Complex Systems. In Duke, R.D. & Kriz, W. (Eds) *Back to the Future of Gaming*. Bielefeld: Bertelsmann,

Meijer, S., & Poelman, R. (2012). Supervisor: a 3D serious game for hazard recognition training in the oil industry. In Bielecki, W., Gandziarowska, J., Pikos, A. & Wardaszko, M. (Eds) *Bonds & Bridges. Facing the Challenges of the Globalizing World with the Use of Simulation and Gaming.* Warsaw: MT Biznes. 209-221

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American psychologist*, 30(10), 955-966

Mihm, J., Loch, C., & Huchzermeier, A. (2003). Problem–solving oscillations in complex engineering projects. *Management Science*, 49(6), 733-750.

Mihm, J., Loch, C. H., Wilkinson, D., & Huberman, B.A. (2010). Hierarchical structure and search in complex organizations. *Management science*, 56(5), 831-848.

Mikulecky, D.C. (2001). The emergence of complexity: science coming of age or science growing old? *Computers & chemistry*, 25(4), 341-348.

Ministerie van Verkeer en Waterstaat [Ministry of Infrastructure and Environment] (2008). Mobiliteitsaanpak [in Dutch]. Den Haag: Ministerie van Verkeer en Waterstaat

Mintzberg, H. (1973). Strategy-making in three modes. *California management review*, 16(2), 44-53.

Mintzberg, H., Raisinghani, D., & Theoret, A. (1976). The structure of" unstructured" decision processes. *Administrative science quarterly*, 246-275.

Mohr, L.B. (1982). *Explaining organizational behavior*. San Francisco: Jossey-Bass

Morton, R.B., & Williams, K.C. (2010). *Experimental political science and the study of causality: From nature to the lab*. Cambridge: Cambridge University Press.

Myers, S., & Marquis, D.G. (1969). *Successful industrial innovations: A study of factors underlying innovation in selected firms.* Washington: National Science Foundation.

Nederlandse Spoorwegen (2016). NS Annual Report 2015. Utrecht: Nederlandse Spoorwegen N.V.

Nelson, R.R. (1992). National innovation systems: a retrospective on a study. *Industrial and corporate change,* 1(2), 347-374.

Nelson, R.R., & Nelson, K. (2002). Technology, institutions, and innovation systems. *Research policy*, 31(2), 265-272.

Nelson, R.R., & Winter, S.G. (1982). *An evolutionary theory of economic change*. Cambridge: Harvard University Press

Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Nikolic, I., & Dijkema, G.P. (2010). On the development of agent-based models for infrastructure evolution. *International journal of critical infrastructures*, 6(2), 148-167.

Nykvist, B., & Whitmarsh, L. (2008). A multi-level analysis of sustainable mobility transitions: Niche development in the UK and Sweden. *Technological forecasting and social change*, 75(9), 1373-1387.

Ordover, J. A., Pittman, R.W., & Clyde, P. (1994). Competition policy for natural monopolies in a developing market economy. *Economics of Transition*, 2(3), 317-343.

Orlikowski, W.J. (2002). Knowing in practice: Enacting a collective capability in distributed organizing. *Organization science*, 13(3), 249-273.

Pavitt, K. (1984). Sectoral patterns of technical change: towards a taxonomy and a theory. *Research policy*, 13(6), 343-373.

Perry, J.L., & Rainey, H.G. (1988). The public-private distinction in organization theory: A critique and research strategy. *Academy of management review*, 13(2), 182-201.

Peters, V.A., Vissers, G.A., & Heijne, G. (1998). The validity of games. *Simulation & Gaming*, 29(1), 20-30.

Peters, V.A., & Vissers, G.A. (2004). A simple classification model for debriefing simulation games. *Simulation & Gaming*, 35(1), 70-84.

Pettigrew, A.M. (1990). Longitudinal field research on change: Theory and practice. *Organization science*, 1(3), 267-292.

Pettigrew, A.M. (1992). The character and significance of strategy process research. *Strategic management journal*, 13(S2), 5-16.

Pettigrew, A.M., Woodman, R.W., & Cameron, K.S. (2001). Studying organizational change and development: Challenges for future research. *Academy of management journa*l, 44(4), 697-713.

Pinch, T.J., & Bijker, W.E. (1984). The social construction of facts and artefacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social studies of science*, 14(3), 399-441.

Pisano, G.P. (1990). The R&D boundaries of the firm: an empirical analysis. *Administrative Science Quarterly* 35(1), 153-176.

Pondy, L.R. (1967). Organizational conflict: Concepts and models. *Administrative science quarterly* 12(2), 296-320.

Poole, M.S., & Van de Ven, A.H. (1989). Using paradox to build management and organization theories. *Academy of management review*, 14(4), 562-578.

Poole, M.S., Van de Ven, A.H., Dooley, K., & Holmes, M. E. (2000). *Organizational change and innovation processes: Theory and methods for research*. Oxford University Press.

Price, W.J., & Bass, L.W. (1969). Scientific research and the innovative process. *Science*, 164(3881), 802-806.

Profillidis, V.A. (2016). *Railway Management and Engineering,* London: Ashgate

ProRail (2016). ProRail Jaarverslag 2015 (in Dutch). Utrecht: ProRail

Puffert, D.J. (2002). Path dependence in spatial networks: the standardization of railway track gauge. *Explorations in Economic History*, 39(3), 282-314.

Qudrat-Ullah, H. (2007). Debriefing can reduce misperceptions of feedback: The case of renewable resource management. *Simulation & Gaming*, 38(3), 382-397.

Ragothama, J. & Meijer, S.A. (2018) Rigor in gaming for design: Conditions for transfer between game and reality. *Simulation & Gaming*. 49(3), 1-17.

Raser, J.C. (1969). *Simulations and society: An exploration of scientific gaming*. Boston: Allyn & Bacon.

Raven, R. (2007). Niche accumulation and hybridisation strategies in transition processes towards a sustainable energy system: An assessment of differences and pitfalls. *Energy Policy*, 35(4), 2390-2400.

Raven, R., Schot, J., & Berkhout, F. (2012). Space and scale in socio-technical transitions. *Environmental Innovation and Societal Transitions*, 4, 63-78.

Reich, Y., Konda, S.L., Monarch, I.A., Levy, S.N., & Subrahmanian, E. (1996). Varieties and issues of participation and design. *Design Studies*, 17(2), 165-180.

Reitsma, R., Zigurs, I., Lewis, C., Wilson, V., & Sloane, A. (1996). Experiment with simulation models in water-resources negotiations. *Journal of water resources planning and management*, *122*(1), 64-70.

Rescher, N. (1996). *Process metaphysics: An introduction to process philosophy*. Albany: SUNY Press.

Rhyne, R.F. (1972). Communicating holistic insights. *Fields Within Fields-Within Fields*, 5, 93-104.

Rip, A., Kemp, R., 1998. Technological change. In: Rayner, S., Malone, E.L. (Eds), *Human Choice and Climate Change*. Columbus: Battelle Press 327–399

Rivkin, J.W., & Siggelkow, N. (2002). Organizational sticking points on NK landscapes. *Complexity*, 7(5), 31-43.

Rivkin, J.W., & Siggelkow, N. (2003). Balancing search and stability: Interdependencies among elements of organizational design. *Management Science*, 49(3), 290-311.

Rosenberg, N. (1976). *Perspectives on technology*. Cambridge: Cambridge University Press

Rosenkopf, L., & Almeida, P. (2003). Overcoming local search through alliances and mobility. *Management science*, 49(6), 751-766.

Rotmans, J., Kemp, R., & Van Asselt, M. (2001). More evolution than revolution: transition management in public policy. *Foresight*, 3(1), 15-31.

Rotmans, J., & Loorbach, D. (2009). Complexity and transition management. *Journal of Industrial Ecology*, 13(2), 184-196.

Ruohomäki, V. (2003). Simulation gaming for organizational development. *Simulation & Gaming*, 34(4), 531-549.

Ryan, T. (2000). The role of simulation gaming in policy-making. *Systems Research and Behavioral Science*, 17(4), 359.

Rycroft, R.W., & Kash, D.E. (1994). Complex technology and community: Implications for policy and social science. *Research Policy*, 23(6), 613-626.

Sahal, D. (1981). Alternative conceptions of technology. *Research policy*, 10(1), 2-24.

Salaman, G., & Storey, J. (2002). Managers' theories about the process of innovation. *Journal of Management studies*, 39(2), 147-165.

Sargent, R.G. (2005). Verification and Validation of Simulation Models. In Kuhl, M.E., Steiger, N.M., Armstrong, F.B. & Jones, J.A. (Eds.) *Proceedings of the 2005 Winter Simulation Conference*, Piscataway: IEEE. 130-143.

Saviotti, P.P., & Metcalfe, J.S. (1984). A theoretical approach to the construction of technological output indicators. *Research Policy*, 13(3), 141-151.

Schmookler, J. (1962). Economic sources of inventive activity. *The Journal of Economic History*, 22(01), 1-20.

Schot, J.W. (1992). Constructive technology assessment and technology dynamics: the case of clean technologies. *Science, Technology & Human Values*, 17(1), 36-56.

Schumpeter, J.A. (1934). *The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle*. New Brunswick: Transaction publishers.

Schumpeter, J.A. (1950). 'The process of creative destruction', in J. Schumpeter (ed.). Capitalism, Socialism and Democracy, Third Edition, London: Allen and Unwin.

Sewell, W.H. (1992). Introduction: Narratives and social identities. *Social Science History*, 16(03), 479-488.

Shove, E., & Walker, G. (2007). CAUTION! Transitions ahead: politics, practice, and sustainable transition management. *Environment and Planning A*, 39(4), 763-770.

Simon, H.A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69(1), 99-118.

Simon, H.A. (1956). Rational choice and the structure of the environment. *Psychological review*, 63(2), 129.

Simon, H.A. (1959). Theories of decision-making in economics and behavioral science. *The American economic review*, 49(3), 253-283.

Simon, H.A. (1969). *The sciences of the artificial*. Cambridge, MA.

Sims, R.R. (2002). Debriefing experiential learning exercises in ethics education. *Teaching Business Ethics*, 6(2), 179-197.

Smith, A., Voß, J.P., & Grin, J. (2010). Innovation studies and sustainability transitions: The allure of the multi-level perspective and its challenges. *Research policy*, 39(4), 435-448.

Smith, A., & Kern, F. (2009). The transitions storyline in Dutch environmental policy. *Environmental Politics*, 18(1), 78-98.

Smith, R. P., & Eppinger, S. D. (1997). A predictive model of sequential iteration in engineering design. *Management Science*, *43*(8), 1104-1120.

Sommer, S.C., & Loch, C.H. (2004). Selectionism and learning in projects with complexity and unforeseeable uncertainty. *Management science*, 50(10), 1334-1347.

Stacey, R.D. (1995). The science of complexity: An alternative perspective for strategic change processes. *Strategic management journal*, 16(6), 477-495.

Staw, B.M. (1981). The escalation of commitment to a course of action. *Academy of management Review*, 6(4), 577-587.

Steinbruner, J.D. (1974) The cybernetic theory of decision. Princeton: Princeton University Press.

Sterman, J.D. (1987). Testing behavioral simulation models by direct experiment. *Management Science*, 33(12), 1572-1592.

Stevenson, H.H. and J.C. Jarillo (1990). 'A paradigm of entrepreneurship: Entrepreneurial management', *Strategic Management Journal*, 11, 17-27.

Strauss, A.L., Corbin, J.M. (1990) *Basics of qualitative research: grounded theory procedures and techniques*, Thousand Oaks: Sage

Suarez, F.F., & Oliva, R. (2005). Environmental change and organizational transformation. *Industrial and Corporate Change*, 14(6), 1017-1041.

Subrahmanian, E., Monarch, I., Konda, S., Granger, H., Milliken, R., & Westerberg, A. (2003). Boundary objects and prototypes at the interfaces of engineering design. *Computer Supported Cooperative Work (CSCW)*, 12(2), 185-203.

Subrahmanian, E., Reich, Y., Smolders, F., & Meijer, S.A. (2011). Designing: Insights from weaving theories of cognition and design theories. Paper presented at the 11th International Conference of Engineering Design, 15–18 August 2011, Kungens Lyngby: Technical University of Denmark

Suurs, R.A., & Hekkert, M.P. (2009). Cumulative causation in the formation of a technological innovation system: The case of biofuels in the Netherlands. *Technological Forecasting and Social Change*, 76(8), 1003-1020.

Taylor, F.W. (1911). *Shop management*. McGraw-Hill.

Teece, D.J. (1996). Firm organization, industrial structure, and technological innovation. *Journal of economic behavior & organization*, 31(2), 193-224.

Koppenjan, J.F.M. & Termeer. C.J.A.M. (1997). Managing perceptions in networks, in Kickert W.J.M, Klijn, E.H. & Koppenjan, J.F.M. (Eds.) *Managing complex networks*. London: Sage. 79-97.

Taylor, J. & Levitt, R. (2004). Understanding and managing systemic innovation in project-based industries, in Slevin, D., Cleland, D., & Pinto, J. (Eds.) *Innovations: project management research*. Newton Square: Project Management Institute. 83-99.

Thiagarajan, S. (1992). Using games for debriefing. *Simulation & gaming*, 23(2), 161-173.

Thietart, R.A., & Forgues, B. (1995). Chaos theory and organization. *Organization science*, 6(1), 19-31.

Thissen, W.A., & Herder, P.M. (2003). Critical infrastructures: Challenges for systems engineering, *Systems, Man and Cybernetics* 2, 2042-2047

Thomke, S.H. (1998). Managing experimentation in the design of new products. *Management Science*, 44(6), 743-762.
Thomke, S.H. (2001). Enlightened experimentation. The new imperative for innovation. *Harvard Business Review*, 79(2), 66-75.
Tidd, J. (1997). Complexity, networks & learning: Integrative themes for research on innovation management. *International Journal of Innovation Management*, 1(01), 1-21.
Trist, E.L. (1981). The evolution of socio-technical systems. *Occasional paper*, 2, 1981.
Trist, E.L., & Bamforth, K.W. (1951). Some social and psychological consequences of the Longwall Method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. *Human relations*, 4(1), 3-38.
Tsoukas, H., & Hatch, M.J. (2001). Complex thinking, complex practice: The case for a narrative approach to organizational complexity. *Human Relations*, 54(8), 979-1013.
Tsuchiya, S. (1998). Simulation/gaming as an essential enabler of organizational change. *Simulation & Gaming*, 29(4), 400-408.
Tushman, M.L., & Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative science quarterly*, 35(4), 439-465.
Tushman, M.L., & Rosenkopf, L. (1992). Organizational determinants of technological-change-toward a sociology of technological evolution. *Research in organizational behavior*, 14, 311-347.
Tversky, A., & Kahneman, D. (1975). Judgment under uncertainty: Heuristics and biases. In Wendt, D. & Vlek, C. (Eds) *Utility, probability, and human decision making*. Amsterdam: Springer Netherlands. 141-162
Utterback, J.M. (1974). Innovation in industry and the diffusion of technology. *Science*, 183(4125), 620-626.
Van Bree, B., Verbong, G.P., & Kramer, G.J. (2010). A multi-level perspective on the introduction of hydrogen and battery-electric vehicles. *Technological Forecasting and Social Change*, 77(4), 529-540.
Van Buuren, A., & Edelenbos, J. (2004). Conflicting knowledge: Why is joint knowledge production such a problem? *Science and public policy*, 31(4), 289-299.
Van de Velde, D. (2000). Dutch and Japanese railway reforms and exchanges. *Japan Railway & Transport Review*, 24, 10-16.
Van de Velde, D.M., & Röntgen, E.F. (2009). Railway separation European diversity. In Auger, J., Bouma, J.J. & Künneke, R. (Eds.) *Internationalization of Infrastructures*. Delft: Economics of Infrastructures. 211-230
Van de Ven, A.H. (1986). Central problems in the management of innovation. *Management science*, 32(5), 590-607.
Van de Ven, A.H. (1992). Suggestions for studying strategy process: A research note. *Strategic management journal*, 13(5), 169-188.
Van de Ven, A.H., & Poole, M.S. (1995). Explaining development and change in organizations. *Academy of management review*, 20(3), 510-540.
Van de Ven, A.H., & Polley, D.E. (1992). Learning while innovating. *Organization Science*, 3(1), 92-116.
Van de Ven, A.H., Polley, D.E., Garud, R., & Venkataraman, S. (1999). *The innovation journey*. New York: Oxford University Press
Van den Brink, W.P., & Mellenbergh, G.J. (1998). *Testleer en testconstructie* [test methodology and construction]. Amsterdam: Boom Koninklijke Uitgevers.
Van den Hoogen, J., Lo, J.C. & Meijer, S.A. (2014). Debriefing in Gaming Simulation for Research: Opening the Black Box of the Non-trivial Machine to Assess Validity and Reliability. In Tolk, A., Diallo, S.D., Ryzhov, I.O., Yilmaz L., Buckley, S. & Miller, J.A. (Eds.) *Proceedings of the 2014 Winter Simulation Conference*, Piscataway: IEEE. 3503-3516
Van den Hoogen, J., Lo, J.C., & Meijer, S.A. (2014). The debriefing of research games: A structured approach for the validation of gaming simulation outcomes. In Kriz, W.C. (Ed.) *The shift from teaching to learning*, Bielefeld: W. Bertelsmann. 88-99.
Van den Hoogen, J., Lo, J., & Meijer, S. (2016). Debriefing research games: context, substance and method. *Simulation & Gaming*, 47(3), 368-388.
Van den Hoogen, J., & Meijer, S. A. (2012). Deciding on innovation at a railway network operator: A grounded theory approach. In *CESUN 2012: 3rd International Engineering Systems Symposium, Delft University of Technology, The Netherlands, 18-20 June 2012*.
Van den Hoogen, J., & Meijer, S.A. (2015). Lessons on the design of gaming simulations for convergence and divergence in volatile innovation environments. In Yilmaz, L., Chan, W.K.V., Moon, I., Roeder, T.M.K., Macal, C. & Rossetti, M.D. (Eds) *Proceedings of the 2015 Winter Simulation Conference*, Piscataway: IEEE. 1044-1055
Van den Hoogen, J., & Meijer, S. (2015). Gaming and simulation for railway innovation: a case study of the Dutch railway system. *Simulation & Gaming*, *46*(5), 489-511.
Van den Hoogen, J., & Meijer, S. (2016). Gaming simulation hybrids in the railway domain: how games impact the volatility of innovation processes. In Kaneda, T., Kanegae, H., Toyoda, Y., Rizzi, P. (Eds.) *Simulation and Gaming in the Network Society. Translational Systems Sciences* (vol. 9). Singapore: Springer. 291-307
Van Der Meij, H., Leemkuil, H., & Li, J.L. (2013). Does individual or collaborative self-debriefing better enhance learning from games? *Computers in human behavior*, 29(6), 2471-2479.
Van Ments, M.V. (1983). *The effective use of role-play: A handbook for teachers and trainers*. London: NP Cogan Page.
Veenendaal, G. (2004). *Spoorwegen in Nederland*. Amsterdam: Boom

Verbong, G., & Geels, F. (2007). The ongoing energy transition: lessons from a socio-technical, multi-level analysis of the Dutch electricity system (1960–2004). *Energy policy*, 35(2), 1025-1037.

Verganti, R. (2008). Design, meanings, and radical innovation: A metamodel and a research agenda. *Journal of product innovation management*, 25(5), 436-456.

Vissers, G., Heyne, G., Peters, V.A., & Geurts, J. (2001). The validity of laboratory research in social and behavioral science. *Quality and Quantity*, 35(2), 129-145.

Von Bertalanffy, L. (1968). *General system theory*. New York: Springer

Von Foerster, H. (1984). *Principles of self-organization—in a socio-managerial context. In Ulrich H. & Probst, G.J.B. (Eds) Self-organization and management of social systems*. Berlin: Springer. 2-24

Von Hippel, E. (1986). Lead users: a source of novel product concepts. *Management science*, 32(7), 791-805.

Von Neumann, J., & Burks, A. W. (1966). Theory of self-reproducing automata. *IEEE Transactions on Neural Networks*, 5(1), 3-14.

Voß, J.P., Smith, A., & Grin, J. (2009). Designing long-term policy: rethinking transition management. *Policy sciences*, 42(4), 275-302.

Walker, W. (2000). Entrapment in large technology systems: Institutional commitments and power relations. *Research Policy*, 29(7-8), 833–46

Weber, M. (1949) *On the methodology of the social sciences*. Free Press

Weber, K.M., & Rohrracher, H. (2012). Legitimizing research, technology and innovation policies for transformative change: Combining insights from innovation systems and multi-level perspective in a comprehensive 'failures' framework. *Research Policy*, 41(6), 1037-1047.

Weckowicz Thaddus, E. (1989). „Ludwig von Bertalanffy (1901–1972): A Pioneer of General Systems Theory" (No. 2). CSR Working Paper.

Weijnen, M.P.C., Bouwmans, I. (2006). Innovation in networked infrastructures: coping with complexity, *International Journal of Critical Infrastructures*, 2, (2-3) 121-132

Wenzler, I., & Chartier, D. (1999). Why do we bother with games and simulations: An organizational learning perspective. *Simulation & Gaming*, 30(3), 375-384.

Westerberg, A.W., Subrahmainan, E., Reich, Y., & Konda, S. (1997). Designing the process design process. *Computers & chemical engineering*, 21, 1-9

Wiener, N. (1948). *Cybernetics*. Paris: Hermann.

Williamson, O.E. (1973). Markets and hierarchies. *American Economic Review*, 63(2) 316-325

Wilson, J.R., Farrington-Darby, T., Cox, G., Bye, R., & Hockey, G.R.J. (2007). The railway as a socio-technical system: human factors at the heart of successful rail engineering. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 221(1), 101-115.

Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: University of Chicago Press.

Wolfe, R.A. (1994). Organizational innovation: Review, critique and suggested research directions. *Journal of management studies*, 31(3), 405-431.

Wooldridge, M., & Jennings, N.R. (1995). Intelligent agents: Theory and practice. *The knowledge engineering review,* 10(02), 115-152.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the sixth international congress of genetics*, Vol. 1, 356-366

Yates, J.F., Jagacinski, C.M., & Faber, M.D. (1978). Evaluation of partially described multiattribute options. *Organizational Behavior and Human Performance*, 21(2), 240-251.

Yin, R.K. (2013). *Case study research: Design and methods*. Thousand Oaks: Sage publications.

Zechmeister, J.S., Shaughnessy, J.J., & Zechmeister, E.B. (2001). *Essentials of research methods in psychology*. New York: McGraw-Hill Higher Education.

Summary

In 2009 ProRail, the Dutch railway infrastructure manager started the use of gaming simulation to support its innovation processes. The organization found that innovations became more systemic and, railways being sociotechnical systems, increasingly involved both changes to technology and human behavior. Subsequently, the organization deemed gaming simulation a valuable addition to existing computer simulations. Such gaming simulations are experiments with models of a system, where human players become part of the simulation. Gaming simulation would for instance allow the organization to experiment with different railway infrastructure layouts around stations and see the effects on network resilience. This is because in this very example human behavior, e.g. in the form of traffic controllers rerouting trains, plays a crucial role. From 2009 onwards a range of gaming simulations have been designed and employed for similar purposes in the Dutch railway sector

Currently however, both practitioners and scholars have built up limited understanding of the use of gaming simulation for innovation processes in sociotechnical systems such as the railways. Firstly, this has to do with the main applications of the tool. Gaming simulation has historically been mostly used for training and education purposes or for policy-making exercises. Secondly, innovation processes are relatively rare in inert sociotechnical systems, especially innovations that we define as systemic: collections of a varied set of innovations that in their conjunction radically change the system. A poor understanding of both causes a problem. This is because it not only remains unknown to what extent gaming simulation can support innovation processes, but also what this support constitutes in the first place. Not knowing the desired functionality of games then renders any design of such games more of an art rather than a craft.

This thesis builds upon the assertion that, according to Klabbers (2003; 2006), the design of gaming simulation needs to closely follow the design of the process in which it is embedded. Games for innovation processes will be significantly different from games for policy-making and training. Hence, studying the design of games needs to occur in conjunction to the study of the innovation process. In this thesis we therefore firstly studied systemic innovation processes in the railway sector independently. In studying innovation processes we adhered to the notion of Poole and Van de Ven (1989) that such processes consist of local mechanisms invoked by intentional actors and resulting emergent patterns. Subsequently this thesis studied how gaming simulation can influence these patterns through these local mechanisms. This thesis thus answered the following main research question.

“What mechanisms play a role in driving a systemic innovation process in the Dutch railway sector and in what ways is gaming simulation able to influence relevant macro-level patterns through these mechanisms?”

The thesis used a qualitative methodology to explore the role gaming simulation can play in systemic innovation processes. For the empirical part, the thesis used the overhaul of Utrecht Central Station as the main case study. This overhaul is where the idea of applying systemically different design principles became manifest. These design principles, originating from Japan, encompass changes to timetabling, routing, safety signaling, track layouts, train and track operator behavior, and rules and roles in case of delays and disruptions. With these principles, some actors within the sector believed that the central node of the Dutch network would be able to accommodate higher volumes of traffic more reliably. The case study is interesting since it poses the first time such systemic change is introduced to the Dutch railway sector. Also, such change inevitably meant different stakeholders with differing professional backgrounds, viewpoints and strategies (and differing willingness to sanction the changes) would become part of the innovation process in different points in time.

The empirical part of this thesis started with a multiple case study to uncover a unique macro-level pattern. We compared the systemic overhaul of Utrecht station with the building of a railway tunnel (incremental) and the introduction of a traffic management system in the UK (radical, but not truly systemic). Here, we saw that systemic change processes do have a unique overall pattern. In stark contrast to usual notions of innovation processes, processes with so-called fuzzy front-ends and increasing stability over time, we found that systemic innovations in the railway sector tend to become more volatile. This occurs as the process moves from exploration to implementation. Using a predesigned analytical framework, we found that dynamics in the design, those who designed it and the rules applicable to the design process were relatively absent during the initial phases. Later in the process these dynamics increased heavily and this pattern appeared to be to some extent dysfunctional.

Manipulating this volatility to make it more evenly spread through time would be a valuable functionality of gaming. For instance, by exploring early on the many ways the innovation makeup can change over time or by early on introducing technical disciplines usually only involved in the process later on, a gaming simulation would be able to front-load volatility and make it more evenly spread throughout time. Also, in volatile times the ability of game to have multiple stakeholders converge on one design, one stakeholder arena and a set of rules would help in making the process more manageable. We translated these features into functionalities of gaming simulation: can these be designed for divergence and for convergence?

After the multiple case study, the thesis continues to zoom in on the Utrecht overhaul case. Using a grounded theory approach, this part of the research built a theoretical framework explaining both the progress over time of the innovation process as well as its tendency to increase in volatility. Based on 25 interviews and a timeline of events and tipping points from 1997 to 2015, the study showed that four mechanisms are crucial in understanding the macro-level pattern. Local stakeholders, knowingly and unknowingly, invoke these mechanisms to progress an innovation. This case study showed that such mechanisms are both constructive (in allowing for progress) and disruptive (in limiting coordination). As local stakeholders invoke the mechanisms for progress in the beginning and later on cause ineffective coordination between stakeholders, these mechanisms in their conjunction and their interplay over time explain why systemic innovation processes tend to increase in volatility rather than settle down: they suppress volatility in earlier stages to allow for progression only to have this volatility surface later on. The four uncovered mechanisms are:

1. Ambiguity
This is the partly intentional process of local stakeholders maintaining an image of an innovation in such a way that inherent dilemmas remain invisible. By doing so, most stakeholders will perceive the innovation as a win-win situation. Ambiguity serves to keep an innovation as an idea alive (or afloat) whilst not yet being implemented. This was found to be highly important since the innovation needed to survive a certain amount of time as an idea because its implementation relied on the serendipitous and timely connection between the innovation idea and a window-of-opportunity. As subsequent mechanisms will show, this serendipitous connection is a unique feature of systemic innovation processes. In this very case, the Japanese design principles were 'discovered' in 1997 but needed the Utrecht renovation project (started more than 10 years later) to latch on. When all the inherent dilemmas were immediately known throughout the sector, it would have made the idea too controversial for innovation champions to pick up this idea. In that case, higher echelons would not have given the innovation champions the organizational resources to further work out the idea.

2. Disentanglement
Disentanglement is the process of operationalizing a systemic innovation from a set of abstract notions on what might constitute this innovation to a fully specified design. In the case under study, we saw that this disentangling involved a deeper study into both the existing Dutch system and the innovation (in this case the Japanese system). The study found that the appearing 'delta' between the current system and the referent system would become the innovation design. It also found that the process of finding this 'delta' was highly path dependent, since it was the result of two mutually impacting processes. In the case study it appeared that what one learns about the referent system strongly depends on what one learns about the current system. Also, during this process the innovation became less systemic as more disciplines became involved in deepening the knowledge on the innovation and these disciplines tended to compartmentalize over time. Disentangling played a significant role in driving forward the innovation process since only an operationalized innovation can be interlocked with another innovation: the so-called

latching-on mentioned at the first mechanism. This interlocking was necessary since the business case for solely the innovation was deemed to be less positive.

3. Interlocking
The actual latching-on of one innovation onto another innovation is called interlocking. Interlocking is a concept already present in existing transition literature. Here it is most often solely seen as a convergent process: multiple innovations coalescing to form a larger innovation. This study showed that in addition to this, interlocking can also cause divergent processes: multiple elements of a single innovation latching-on to different separate innovation projects. This is because for stakeholders interlocking serves two purposes. Firstly, it allowed the innovation to use a window-of-opportunity and made the innovation more implementable. The introduction of Japanese design principles needed the already planned (and financed) renovation of Utrecht Central station and the introduction of timetable-less transport along the Amsterdam – Eindhoven corridor. Using the momentum of these larger projects, the innovation interlocked with these projects and created a tipping point for its own implementation. This highly resembles the convergent interlocking often mentioned in transition literature. This case was also interesting since separate projects solely focused on introducing Japanese design principles at smaller nodes in the network, where hence no interlocking occurred, did not reach full implementation. Secondly, the case showed that during implementation interlocking was an effective way for innovation stakeholders to cope with complexity. We saw that over the course of the innovation process project managers, engineers and designers connected multiple projects to better handle the technical complexity. For example they assumed, for their own infrastructure layout design, that elsewhere in the Netherlands a bridge was to be built, essentially connecting parts of the Utrecht project with a bridge project. It is this linking that creates high volatility later on since the project is no longer shielded from dynamics in other projects.

4. Shielding
Shielding is the process of keeping away regime pressures and pressures of concurrent innovation processes from ones' own innovation process. Shielding enabled stakeholders to design the innovation beyond the initial scope of their responsibility. The innovation champions in this case used design elements far beyond the scope allocated to them by their organizational location and discipline. For instance, in their plan to introduce Japanese design principles they had to include an overhaul of the signaling layout alongside the railway tracks, which is a usual responsibility of the safety departments within the sector. In essence, shielding allows for technical interlocking of projects without the social and institutional interlocking of project teams coordinating their efforts. In the bridge example mentioned earlier it was merely technical interlocking, creating the effect that when the bridge project changed (it got cancelled) the project team involved in the overhaul of Utrecht was not immediately notified.

The four mechanisms point to the serendipitous nature of systemic innovation processes. In addition, the mechanisms all encompass the notion that systemic innovation processes inherently consist of multiple parallel projects that have to interlock over time. Both of these dimensions of systemic innovation processes plead for an entirely different management approach, and subsequently a different conceptualization of the role of decision support tools such as gaming simulation. For instance usual project management principles would overly restrict the innovation in finding windows-of-opportunity. Also, level of knowledge was never a crucial driving force or blocking factor over the course of the innovation process. Hence, a gaming simulation need not simply be a 'knowledge-production-device' such as a classical experiment would be. Nor would it be simply a device to increase levels of collaboration since non-collaboration explained partly the progress of the process in the beginning. If valuable, a gaming simulation needs to enable local stakeholders to more intelligently control the four mechanisms. By doing so, stakeholders can navigate the dilemma between progress and coordination.

The thesis continued with a multiple case study on different applications of gaming simulation in the Dutch and British railway sector. This showed that games indeed could be designed in order to influence effectively the four relevant mechanisms. In general we found that games can alleviate shields, decrease ambiguity and increase interlocking. Games also force innovation stakeholders to more fully disentangle (i.e. operationalize) their innovation since workable models are needed before a game can be played. In addition to that, the study found that during the simulation run players are able to further disentangle an innovation. This means that in-game design is effectively using local knowledge of operators to further workout the details of an innovation. As long as the mechanisms can be seen as constructive, these influences of the use of gaming can be seen as valuable. However as this thesis shows, the mechanisms contain a constructive and disruptive side. Some of the games indeed allowed for too much interlocking and alleviated an ambiguity that could have been constructive. Hence, given the effects on mechanisms, and the mechanisms' link with the macro-level pattern, the value of gaming simulation is highly dependent on the timing of its use and may very well harm an innovation process.

This study then looked at how the studied games impacted the macro-level pattern of volatility. It found that there are two types of gaming simulations: games for convergence and games for divergence. In essence, games for convergence are games that allow stakeholders to determine one single innovation and one single implementation process. These games are valuable in times when the process is highly volatile. In the case of systemic innovations, this is usually in later stages of the process. The final part of the thesis studied games' true ability to allow for such convergence. It found that validity issues play a role and such issues can be tackled in a carefully designed debriefing. However, it also found that games have an inherent tendency to allow for divergence: the exact opposite of convergence. This is the second type of gaming simulation we could distinguish. The bringing together of different stakeholders, the malleability of the game model and experimental design, and the introduction of human game players, all tended to induce exploration. Such exploration is valuable since we had shown that by front-loading volatility, disruptive levels of volatility in later stages of the innovation process could be avoided.

Whether a game is one of these two types depends on certain design parameters. This thesis provided a first step towards these parameters. It found that openness of the game design process and game play, type of game players, the level of detail, the debriefing and the experimental design make a game, mediated by the four mechanisms, either increase (divergent) or decrease (convergent) process volatility. The analysis in this part of the chapter also showed that a divergent function of gaming needs the careful design of the stakeholder arena beforehand. Otherwise, the many insights the game will deliver will not be translated into the innovation process.

This thesis found that the 'active substance' of gaming simulation is the collective envisioning of different *innovations* and potential *processes* in a single environment at a single point of time. In this thesis this is termed as exploring the product (P-)space and exploring the socio-institutional (S- and I-)space respectively. This envisioning enables games to have impacts on mechanisms and patterns and allows it to be of value for innovation processes. This feature can be used to finally arrive at one innovation and one process (termed convergence). In this case, the envisioning of other innovations and processes helps in determining the robustness of the results or help to plan for contingencies. This is valuable because rarely will the innovation, as tested in-game, be the innovation that is implemented in real-life. The feature can also be used to arrive at many innovations and many possible processes (termed divergence). Such exploration would for instance increase the chance the innovation will find valuable interlocking possibilities with parallel innovation processes occurring outside of the game. This is especially relevant since systemic innovations rarely are able to build up momentum themselves, as this thesis showed. Also it increases the chance that actors, disciplines and insights are taken into account early on in the process, thereby front-loading volatility otherwise found in later stages. In both cases however, there needs to be careful consideration of the appropriate actor network to act on the results outside of the game.

The thesis found that this collective envisioning is mostly a cognitive effort of game players and game observers, rather than something automatically realized by a game run itself. This impacts the way gaming simulations are designed and increases the importance of facilitation and debriefing. It namely appeared that game players and observers interacting with the game model and game run invoked the constant asking of 'what-if'-questions. Asking these questions helped in created the exploratory nature of games for divergence or allowed for validity and robustness assessments for games for convergence. Such exploration of eventualities could also be ensured by considering different experimental designs but would always entail running the game many times. Practical concerns, such as resources and game player' availability, limit however the amount of runs. Also the thesis found that during a game run there is little possibility to discuss the innovation process in parallel. There were no instances where a game run allowed for the envisioning of innovations and processes simultaneously. Hence solely executing a game run will not provide the active substance, nor will it ensure gaming simulation having an impact on the process in which it is embedded.

This thesis concluded that it is the debriefing where collective envisioning can take place. Here a facilitator can ensure the collective and transparent collection and assessment of the eventualities considered by game players and observers. In this debriefing phase, stakeholders can share their ideas on how game results could have changed given different starting setups, different innovations or different scenarios. Also such ideas can be actively encouraged and invoked during this phase by a facilitator asking 'what-if'-questions. Based on event-structure analysis methods from the historical and sociological sciences, this thesis provides a set of guidelines for collectively considering these 'what-if'-questions in a structured manner. Also in this phase, stakeholders can collectively decide on the proper actor networks and institutions by which the innovation can be implemented.

This thesis concludes that gaming simulation, as environments where stakeholders can collectively envision a vast set of different innovations and innovation processes, allow for the manipulation of process volatility. Games do so by impacting four relevant mechanisms and effectively enable stakeholders to more intelligently trade-off progress with coordination. This impact is mostly realized by an effective debriefing.

Samenvatting

Vanaf 2009 begon ProRail, de Nederlandse spoorinfrastructuurbeheerder, met het gebruik van spelsimulatie om innovatieprocessen te ondersteunen. De organisatie vond dat innovaties in het spoor, met het spoor als socio-technisch systeem, steeds vaker verandering aan zowel techniek als menselijk gedrag behelsden. Hierdoor achtte zij spelsimulatie als waardevolle toevoeging aan het bestaande instrumentarium van computersimulaties. Zulke spelsimulaties zijn experimenten met modellen van een systeem, waar menselijke spelers onderdeel worden van de simulatie. Spelsimulatie stelt dan bijvoorbeeld de organisatie in staat om te experimenteren met verschillende infrastructuurontwerpen rond een station en de invloed op de veerkracht van een dergelijk ontwerp in geval van grote vertragingen in de dienstregeling. In dit geval speelt menselijk gedrag, zoals treindienstleiders die treinpaden inplannen, een cruciale rol. Vanaf 2009 heeft de organisatie met TU Delft onder het Railway Gaming Suite project verschillende spelsimulaties ontworpen en uitgevoerd.

Het ontwerpen en uitvoeren spelsimulatie in een dergelijke context is echter vrij nieuw. Spelsimulaties worden veelal ingezet voor trainingsdoeleinden of beleidsinterventies maar zelden voor het ontwerpen en testen van technologische innovaties. Bovendien komt het type innovaties dat centraal in dit proefschrift staat vrij beperkt voor in sectoren zoals het spoor. Dit onderzoek richt zich namelijk specifiek op systemische innovaties. Dit zijn innovaties die een groot scala een kleinere veranderingen behelzen en in hun gezamenlijkheid het evolutionaire ontwikkeltraject van een systeem ombuigen. Dit begrip van innovatie neemt in ogenschouw dat historische keuzes sterk bepalen in welke richting een systeem groeit: er is pad-afhankelijkheid. Zoals er ooit gekozen is voor de Qwerty-inrichting van toetsenbord, is de transitie naar een andere inrichting kostbaar, moeilijk en tijdrovend.

Een dergelijke transitie vindt ook plaats in de spoorsector. De ombouw van Utrecht Centraal Station is hier een eerste uitingsvorm van. De onderliggende ontwerpprincipes die bij de ombouw zijn gebruikt, komen namelijk voort uit de waarneming dat een ander systeem dan de Nederlandse, het Japanse spoorsysteem, met dezelfde middelen beter presteert. Ergens in de ontwikkeling van de spoorsystemen heeft dit land gekozen voor eenvoud als leidende filosofie, terwijl het Nederlandse spoor flexibiliteit aanhield. Lange tijd was het Nederlandse spoor hierdoor in staat om onder omstandigheden van niet al te grote benutting van de capaciteit een sterk klantgerichte dienstregeling aan te bieden. Onder stijgende benutting en daarmee gepaard gaande capaciteitsproblemen, blijkt deze historische keuze niet langer in het voordeel te werken.

Dit proefschrift verkent hoe dergelijke transitieprocessen in de spoorsector werken en hoe spelsimulatie hierbij ondersteuning kan verlenen. Dit is sterk ingegeven door de noties uit Klabbers (2009) dat het ontwerpen van spelsimulaties sterk interacteert met het ontwerpen

van het proces ‘daarbuiten’: het daadwerkelijke proces waar een spel invloed op moet hebben. Het onderzoek begint met een verkenning van twee benaderingen op de relatie tussen innovatie en spelsimulatie: een analytisch-wetenschappelijke en een ontwerpwetenschappelijke. De eerste veronderstelt dat innovaties lineair verlopen en dat de rol van spelsimulatie die is van een zuiver experiment. De tweede benadering zegt dat de relatie complexer is omdat processen zelden lineair zijn en omdat spelsimulaties zich niet in een vacuüm afspelen. Om deze innovatieprocessen in het onderzoek verder te ontleden, maakt dit proefschrift gebruik van de waarneming van Poole en Van de Ven (1986) dat innovatieprocessen zowel een microniveau kennen van intentioneel gedrag van actoren, als emergente patronen op macroniveau. In de onderzoeksvraag zijn deze noties verwerkt:

“Welke mechanismen spelen een rol in het voortstuwen van een systemische innovatieproces in de Nederlandse spoorsector en op welke manieren is spelsimulatie in staat relevante patronen op macroniveau te beïnvloeden middels deze mechanismen?”

Het empirische gedeelte van dit proefschrift kent drie delen, waarbij het eerste deel op zoek gaat naar een relevant patroon op macroniveau. Voor dit doel vergelijkt het onderzoek drie innovatieve projecten in de spoorsector waarvan van de voorkant al duidelijk is op welke aspecten de onderliggende innovaties verschillen. Een project was het in zeer korte tijd indienststellen van een nieuwe spoortunnel met ondergronds station in de gemeente Delft. Een tweede project was het invoeren van een complex verkeersleidingsysteem in he Britse spoorsysteem, een radicalere ingreep maar niet volledig systemisch. Het derde project was DSSU, Doorstroomstation Utrecht, het project dat centraal staat in dit proefschrift en uiting gaf aan het ‘Japaniseren’ van het spoor. Het systemische hieraan is dat het onder andere ingrepen in het seinstelsel, spoorontwerp en wissels, bijsturingsregels en dienstregelingen omvat en dat al deze veranderingen in gezamenlijkheid dienen ingevoerd te worden.

De vergelijking van projecten liet zien dat gebruikelijke dynamieken in innovatieprocessen, ‘de fuzzy front-end’ en afzwakkende volatiliteit, zoals verondersteld in de innovatieliteratuur, alleen herkenbaar waren in minder systemische innovatieprojecten. Daarentegen werd zichtbaar bij project DSSU dat de dynamiek in technisch ontwerp, stakeholders en instituties steeds sterker onderhevig waren aan veranderingen naarmate het proces zijn apotheose naderde. Dit heeft, zoals het onderzoek laat zien, sterk te maken met de specifieke architectuur van de onderliggende innovatie: de elementen van de innovatie zijn makkelijk los te koppelen (zoals spoorontwerp en seinstelsels) en kunnen deels al afzonderlijk geïmplementeerd worden. Het onderzoek liet zien dat stakeholders buiten het project bijvoorbeeld enkele ingrepen versneld konden invoeren om aan andere wensen te voldoen.

De stijgende volatiliteit bleek op enkele vlakken dysfunctioneel te zijn en daarmee als emergent patroon een relevante focus te kunnen zijn voor het ontwerpen van spelsimulaties. Een spelsimulatie dient dan volatiliteit naar voren kunnen halen, bijvoorbeeld door al vroeg in het proces stakeholders bij het proces te betrekken die anders pas later hun invloed zouden pakken. Om deze mogelijkheden voor spelsimulaties verder te verkennen heeft het onderzoek zich vervolgens gericht op de onderliggende mechanismen die verklaren waarom dit patroon zich voordoet. Een gevalsstudie naar project DSSU, waarin dieper wordt ingegaan op het handelen van actoren en waarbij in grotere mate contextinvloeden worden meegenomen, heeft deze mechanismen opgeleverd.

In het onderzoek naar DSSU zijn 25 betrokken actoren geïnterviewd, is een tijdlijn opgesteld van 1997 tot 2015 en zijn bepaalde doorslaggevende gebeurtenissen in kaart gebracht. Op basis hiervan kon het onderzoek komen tot een viertal mechanismen die actoren moedwillig of onbewust activeren om een innovatieproces voort te stuwen. De analyse laat zien dat dergelijke mechanismen zowel een functionele als dysfunctionele kant hebben: dit verklaart dat door de persistente inzet van deze mechanismen volatiliteit bij aanvang van een innovatieproces wordt onderdrukt, maar later in hevigere vorm alsnog zich voordoet. De vier mechanismen zijn *ambiguïteit, ontknopen, samenknopen* en *afschermen.* In hun functionele vorm dragen deze mechanismen bij aan de voortgang van een innovatieproces. In hun dysfunctionele vorm gaan ze echter coördinatie tegen. De gevalsstudie laat zien dat dit het grote dilemma is tijdens een dergelijk proces: coördinatie versus voortgang.

Ambiguïteit
Het proces van het niet helder maken van of onbesproken laten van inherente dilemma's die aan een innovatie vastzitten. De ambiguïteit van een innovatie stelt het instaat om langere tijd te 'overleven', ook als idee: het wordt door de sector niet meteen van tafel geveegd omdat meerdere partijen in de innovatie, soms onterecht, een win-win-oplossing zien. Het overleven van een innovatie is gebleken cruciaal te zijn omdat systemische innovaties sterk leunen op het tijdig samengaan met andere projecten. Systemische innovatieprocessen kennen namelijk een sterke serendipiteit doordat ze niet in een vacuüm plaatsvinden.

Ontknopen
Het proces van het operationaliseren van een abstract idee, een kluwen van allerlei met elkaar gerelateerde deeloplossingen tot een specifiek ontwerp. Dit onderzoek liet zien hoe dit proces bestond uit zowel het in onderzoek nemen van het Japanse systeem alsmede het eigen Nederlandse systeem en dat daarmee de 'delta' het eigenlijk ontwerp begon te worden. Opmerkelijk hier is de padafhankelijkheid van dit soort processen: wat men leert over het andere systeem heeft een wederzijdse invloed op wat men leert over het eigen systeem. Ontknopen is een relevante mechanisme omdat enkel een meer ontknoopte innovatie kan samengaan met een ander project en daarmee de kansen voor implementatie sterk vergroot.

Samenknopen

Dit samengaan, of samenknopen, is het combineren van verschillende innovatieprojecten in een programma of het combineren van elementen van een innovatie met andere elementen van andere innovaties. DSSU omvatte bijvoorbeeld het verwijderen van spoorwissels, een ingreep die vanwege de storingsgevoeligheid van wissels, na slechte spoorprestaties gedurende enkele winters, ook onderdeel kon worden van een ander actieprogramma. Dit samenknopen heeft dus niet alleen een convergent karakter, maar kan ook divergent zijn. In het laatste geval valt een innovatie uiteen omdat de elementen die voorheen de innovatie waren nu ingebed in andere projecten afzonderlijk worden ingevoerd. Het diepere onderzoek naar DSSU liet ook zien dat samenknopen een sterk cognitief karakter heeft. Voor stakeholders om om te gaan met de complexiteit van het ontwerpproces werd geregeld het project technisch samengeknoopt met andere projecten. Zo werd bij het ontwerp van Utrecht Centraal op een gegeven moment uitgegaan van een specifieke nieuwe spoorboog elders in het land. Dit maakte de opgave voor Utrecht zelf minder complex, maar ook alleen omdat organisatorisch de projecten niet samengeknoopt werden. Op het moment dat de spoorboog geen doorgang vond, leverde dit weer extra complexiteit (en volatiliteit) op voor het DSSU project.

Afschermen

Innovatieprocessen zijn gebaat bij afscherming van druk vanuit de omgeving, zoals van de staande organisatie of van parallel lopende innovatieprocessen. Op deze manier kan complexiteit gereduceerd en voortgang gerealiseerd worden. Voortgang werd bijvoorbeeld gerealiseerd in het vorige voorbeeld waarbij een spoorboog ver buiten Utrecht onderdeel werd van het project, maar enkel op technisch vlak. Door afscherming waren de stakeholders in staat dit samenknopen niet op organisatorisch vlak ook te laten geschieden: iets wat de zaken alleen maar complexer had gemaakt.

De vier mechanismen laten de serendipiteit zien van systemische innovatieprocessen en tonen het belang aan van verschillende innovatieprojecten die door de tijd heen samengaan en uit elkaar gaan. Deze noties zijn relevant omdat het mogelijk andere management principes verondersteld: als systemische verandering niet langer voortkomt uit een enkel project, maar uit meerdere, worden projectmanagementprincipes niet langer toepasbaar.

Door een verscheidenheid van in de spoorsector uitgevoerde spelsimulaties te onderzoeken is vervolgens in kaart gebracht hoe spelsimulatie deze mechanismen kan beïnvloeden. Het blijkt dat spelsimulaties in staat zijn deze mechanismen te beïnvloeden door specifieke ontwerpkeuzes over de inrichting van het spelontwerpproces en de daadwerkelijke uitvoer. Omdat de mechanismen zowel functioneel als dysfunctioneel kunnen zijn, kan spelsimulatie ook innovatieprocessen tegenwerken: ze kunnen zorgen voor te weinig ambiguïteit als inherente dilemma’s door een spel al te vroeg naar boven komen en ze kunnen leiden tot teveel stakeholders die zicht krijgen op de innovatie en deze met andere innovaties willen samenknopen.

Om te bepalen hoe spelsimulaties dan uiteindelijk doorwerken op het emergente patroon van steeds groter wordende volatiliteit maakt dit onderzoek onderscheid tussen spellen voor divergentie en spellen voor convergentie. Spellen voor convergentie zijn spelsimulaties die zo ontworpen zijn dat ze resulteren in een eindontwerp met daarbij duidelijkheid over de noodzakelijke stakeholders en implementatiestappen. Spellen voor divergentie zijn juist spelsimulaties die uitnodigen tot het exploreren van verschillende ontwerpen en het verkennen van mogelijke stakeholders en implementatiestappen. Dit onderzoek laat zien dat juist in een hoog-volatiele context, waar spellen voor convergentie waardevol zouden kunnen zijn, spellen neigen naar divergentie. Bovendien hebben spelsimulaties zelf een inherente beweging richting divergentie, met name omdat dergelijke exercities de enige momenten zijn dat stakeholders van verschillende delen van de sector, waaronder operationeel personeel dat als speler fungeert, samen komen.

Detailniveau van het spel, het soort speler en het gebruik van echte tijd zijn enkele ontwerpkeuzes die gemaakt kunnen worden om een spel divergent of convergent te laten zijn. De analyse van de werking van spelsimulatie in het tegengaan van volatiliteit later in het proces (convergent) of opwekken ervan eerder in het proces (divergent) toont bovendien aan dat bij het ontwerpen en uitvoeren van spellen voor divergentie, het noodzakelijk is van te voren de juiste stakeholders te betrekken. Dit om de waardevolle uitkomsten van een spel te laten landen bij diegene die kunnen handelen.

Dit onderzoek concludeert dat de waarde van spelsimulatie voor systemische innovatieprocessen ligt in de mogelijkheid om in gezamenlijkheid een verscheidenheid van innovaties, actoren en innovatieprocessen te verkennen. Dit verkennen, is gebleken, is vrij cognitief: het is een exercitie die spelers en waarnemers van het spel samen doen tijdens een spel, maar in grotere mate tijdens de debriefing. Daarmee wordt de cruciale rol van debriefing nu ook bevestigd voor spelsimulaties voor innovatieprocessen, naast die voor trainingsdoeleinden en beleidsinterventies. Dit onderzoek presenteert een methode om de debriefing nog beter in te zetten voor dit cognitieve verkenningsproces. Door dit proces is spelsimulatie in staat via de vier mechanismen het emergente patroon van volatiliteit te beïnvloeden.

Het onderzoek geeft hiermee een nieuwe invulling aan de waarde van spelsimulatie in technische, en systemische, innovatieprocessen. In tegenstelling tot veronderstellingen vanuit een analytisch-wetenschappelijk perspectief, ligt de waarde van spelsimulatie niet uitsluitend in de mogelijkheid haar in te zetten als zuiver experiment. Dit laat namelijk onverlet dat een inhoudelijke onzekerheid of kennistekort zelden een cruciale rol speelt in de dynamieken van een dergelijk proces. Wel geeft dit proefschrift weer dat thema's vanuit dit perspectief, zoals validiteitsoverwegingen, een rol spelen in het versterken van het effect van spelsimulatie, maar het effect zelf niet omschrijven. Een meer bruikbare conceptie van het begrip 'waarde' geeft dit proefschrift: het opwekken of tegengaan van volatiliteit door middel van de vier mechanismen.

In relatie tot de literatuur over innovatieprocessen en transities in socio-technische systemen voegt dit onderzoek een aantal interessante noties toe. Zo is sterk bepalend gebleken hoe de architectuur van een innovatie zich ontwikkeld gedurende het proces omdat het stakeholders wel of niet in staat stelt invloed uit te oefenen op het proces. De rol van het technische artefact, of innovatie, is tot nu toe altijd onderbelicht gebleven. Bovendien laat dit onderzoek zien hoe processen van samenknopen sterk afhangen van de mate waarin een innovatie onder de radar kan overleven, ook als idee. Met name in de transitieliteratuur is samenknopen een bekend gegeven, maar de noodzaak van het ambigu houden van een innovatie om dit samenknopen mogelijk te maken is een interessante toevoeging. Daarbovenop toont dit onderzoek dat samenknopen kan leiden tot divergerende innovatieprocessen in plaats van uitsluitend convergerende processen. In het door actoren bouwen van momentum gedurende een transitie, kern van veel van de normatieve modellen in de transitieliteratuur, is divergentie zeker een proces dat hier sterke invloed op kan hebben.