DRAFT VERSION DECEMBER 2016

From one to many: generalisation and evidence in failure analysis

Inge De Bal

Centre for Logic and Philosophy of Science
Ghent University (UGent)
Blandijnberg 2, B-9000 Gent, Belgium
Inge.DeBal@UGent.be

Keywords
Philosophy of engineering; mechanisms; failure analysis; generalisation; evidence; Daniel Steel; Nancy Cartwright; case studies; design perspective

Abstract
In this paper, I use cases and reasoning from failure analysis (a part of engineering science which deals with artefact failure and its causes) to draw attention to a relatively unstudied problem of knowledge generalisation: when we are focusing on *creating* new things; designing new artefacts and technologies. Using three cases from failure analysis practice, I present a two-fold mechanism-based procedure to determine when generalisations to non-existing artefacts are warranted. This procedure builds on (1) Cartwright's notion of capacities (2) literature on mechanisms and (3) Steel's comparative process tracing, developed for the biomedical sciences. I will show that, while they provide guidance, these literatures and concepts are not enough to grasp how we use information from current artefacts and failures to create new things – we will need a lot more specific information and adequate ways to present it. The account developed in this paper is relevant for both philosophers and failure analysts. For philosophers, it can provide input for a theory of evidence. For failure analysts, it allows them to present stronger arguments for their recommendations by making the required evidence explicit. My account can furthermore provide inspiration for similar inferences in other innovation contexts such as pharmacology.

Acknowledgments
I would like to thank the Fonds voor Wetenschappelijk Onderzoek for supporting this research.
I am grateful for helpful comments by participants of MuST9 and SPSP 2016. The work has been significantly improved due to suggestions from Erik Weber and Phyllis Illari. Comments by anonymous referees proved particularly rewarding.

From one to many: generalisation and evidence in failure analysis

1. Introduction

When an artefact breaks down, specialized engineers called failure analysts study the specific circumstances that led to this failure. For instance, in "Creep failure of a spray drier", Paul Carter investigates the collapse of a specific spray drier[1] which had been in service for nearly 20 years (2001, 73). This article was reprinted in *Failure Analysis Case Studies II*, a collection of "40 case studies describing the analysis of real engineering failures which have been selected from volumes 4, 5 and 6 of Engineering Failure Analysis[2]" (Jones 2001, v). In the preface, the editor comments on the previous edition:

> The book has proved to be a sought-after and widely used source of reference material to help people avoid or analyse engineering failures, design and manufacture for greater safety and economy, and assess operating, maintenance and fitness-for-purpose procedures. (ibid.)

Although failure analysts start from specific case studies, they do not simply want to explain what happened in this one situation. They also seek knowledge to prevent similar problems in the future. This is expressed in the quote above and also aptly stated by Petroski:

> When failures do occur, engineers necessarily want to learn the causes. Understanding of the reason for repeated failures […] typically leads to a redesigned product. (2001, 13)

In other words, failure analysts look for ways to use the knowledge about causal relations in one specific situation, to draw conclusions regarding causal relations in other situations. These situations range from other instances of the same artefact, over similar artefacts, and even to very different artefacts. One of their goals is furthermore to find ways to alter designs. Their analysis is thus thought to be useful for

> (1) understanding (failure of) existing artefacts
>
> (2) altering practices of use of these existing artefacts
>
> (3) designing new artefacts not yet in existence.

In this paper, I attempt to make sense of how this can be done. This seems to be an instance of a longstanding philosophical problem regarding generalisation of knowledge from one particular instance or local domain, to other instances or domains. This problem has occurred under many different names and in slightly different forms, including 'induction', 'extrapolation' and 'external validity'. Arguably, these problems and debates are similar in the sense that they deal with the question of how to generalize knowledge. I will investigate different types of generalisation as they occur in failure analysis and what evidence is given/needed for them. Because of the focus on design, the generalisations in failure analysis differ from the more classic examples. The design perspective sets them apart. Understanding these generalisations can deepen our philosophical understanding of different ways generalisation problems occur and how to solve them. In section 2, I first discuss induction and extrapolation in more detail. I then spend some attention to the notion of design and

[1] A spray dries is an artefact often used in mines to dry liquid or slurry fast by means of hot gas.
[2] Engineering Failure Analysis is a journal which "publishes research papers describing the analysis of engineering failures and related studies" (Elsevier 2016).

clarify what I mean with 'design perspective'. In section 3, I present three examples from failure analysis practice. They will serve as case studies throughout the paper. I will flesh out three distinct types of inference: one that looks like induction, one that looks like extrapolation and one that is still different. I will argue that none of the examples are 'pure' instances of classical generalisations, because they involve artefacts-to-be-designed. They will further specify what is meant with the design perspective of generalisations. Throughout the paper I develop a framework to analyse these inferences. It builds on existing philosophical literature, but I make suitable adaptations to capture the implications for non-existing artefacts. In order to analyse the aforementioned inferences, I first use Cartwright's notion of capacities to present the underlying causal claims and their domains in a standard format. They will allow me to capture the probabilistic nature and locality of the causal claims, while accounting for the stability required for generalisations. Using this standard format, I will clarify the (implicit) inference steps analysts make in their causal generalising reasoning. This will shed light on the evidence required to warrant these steps. Because of the focus on design, we will need a lot of specific information to ensure that recommendations prescribe warranted changes to designs. This is the topic of sections 5 and 6. There, I develop my mechanism-based account of the evidence needed to warrant the aforementioned inference steps. It builds on Steel's framework regarding extrapolation in the biomedical sciences. Because the inferences I am concerned with are not strict extrapolations, I will adapt Steel's account, building on sections 3 and 4. I first determine a mechanism-based criterion of similarity for artefacts. Then, I define a mechanism-based heuristic to determine when generalisations are warranted in failure analysis. This will also create a clear picture of the required evidence for such generalisations. To finish, I look back on the tools I used and elaborate why they needed adjusting and developing to guide us in building new artefacts. I also reflect on the nature of these additions to understand what was missing and how this can be of help in other domains.

Most of the philosophical tools and literature I use in this paper originated from philosophical interest in the biomedical and social sciences. In this literature, technological contexts are often contrasted with the biomedical and social cases the authors focus on. Cartwright and Hardie, in their book entitled "Evidence Based Policy"[3], e.g. write

> It would be nice if social policy were like a battery. Everything necessary for it to create a current is locked inside the casing; the environment it is to be put to work in is both structured and delimited, like a flashlight or a radio; and there are clear instructions for how it is to be implemented—"Put the end marked *+* here." But for social policies, the requisite scientific and technological knowledge and know- how is often missing. (2012, 91-92)

There is a lot of truth to this quote. Yet as will become clear from the examples in section 3, making an artefact like a battery work is not as straightforward as Cartwright and Hardie suggest – let alone adapting it in a successful way.

[3] I thank an anonymous reviewer for suggesting this. I will refer to this book several more times. The goal of their book is comparable to the goal of my paper, but they focus on the social sciences. As will become clear later in the paper, because of my focus on design, their strategies for going from "it works somewhere to it will work here" (2012, 15) are not very useful in the context of this paper.

2. Generalisation in failure analysis

2.1 Generalisation problems

In order to frame the current paper, I briefly present an overview of several of the aforementioned generalisation problems and related philosophical concepts. Taking a closer look at these debates will show that they all share a common concern: the question of knowledge generalization. More importantly, these debates show that there is no clear consensus about what this question entails or how to solve it. This overview will allow me to develop the discussion of failure analysis more efficiently and show exactly what studying failure cases can teach us.

I will start with induction. The problem of induction has a long-standing history in philosophy. Hume is generally considered the first to draw attention to it:

> As to past *Experience*, it can be allowed to give *direct* and *certain* information of those precise objects only […] which fell under its cognizance: but why this experience should be extended to future times, and to other objects, which for aught we know, may be only in appearance similar; this is the main question (1748, part 4, §29)

Russell defines it as one of the major problems in philosophy:

> It must be known to us that the existence of some one sort of thing, A, is a sign of the existence of some other sort of thing, B, either at the same time as A or at some earlier or later time […]. The question we have now to consider is whether such an extension is possible, and if so, how it is effected. (1912, 39)

These concerns have not gone unstudied. Philosophers like Mill (1843), Peirce (1883) and Carnap (1950) spent significant philosophical attention on the problem of induction. Gradually, the problem took on different forms, like the paradox of the ravens (Hempel 1945) and Goodman's new riddle of induction (1955). Others, like Maher, attempt to model induction formally (1999). The main question, however, still underlies all these on-going debates: the definition of induction is not agreed upon (Vickers 2016), nor has the problem been solved to everyone's satisfaction (Norton 2003). But regardless of the specific definition of (the problem of) induction, these enquiries all engage with the question of how we can justifiably *generalise* knowledge of observed events to unobserved ones.

Another generalisation problem can be found in philosophy of the biomedical and social sciences, specifically in debates regarding *extrapolation*. Steel (2007) introduces the problem of extrapolation as follows:

> Imagine that a chemical […] has been found to be carcinogenic if administered […] in rats, and the question is whether it is also a carcinogen in humans. This is an example of extrapolation: given some knowledge of the causal relationship between X and Y in a base population, we want to infer something about the […]target population. (78)

Where the problem of induction posed the question in terms of observed and unobserved events, the extrapolation problem focuses on how to generalize knowledge between different populations. Illari and Russo (2014) argue that the widely discussed extrapolation[4] problem is important for both

[4] Illari and Russo also connect it to another topic, namely the problem of external validity (2014, p.18). External validity has been discussed by many philosophers, both formally and informally see e.g. Pearl and Bareinboim (2014), Jimenez-Buedo and Miller (2010), Guala (2005). It bears striking resemblance to the problem of

observational and experimental studies (p.48). I do not claim to have presented a complete overview of philosophical literature regarding induction and extrapolation. Yet I hope to have shown that generalization of knowledge is a significant problem that underlies multiple debates, including debates on induction and extrapolation.

2.2 The design-perspective

In this paper, I use cases and reasoning from failure analysis to draw attention to another way in which the generalisation problem arises: when we are focusing on creating new things. Both induction and extrapolation deal with events or populations *that already exist*. To be more precise, induction and extrapolation focus on whether our current knowledge generalizes to already existing things that we have not studied[5]. These are different questions than whether our current knowledge can help us *to create something new*. As Von Karman famously pointed out, creating new things is central to engineering:

> Scientists discover the world that exists; engineers create the world that never was. (Bucciarelli 2003, 1)[6]

This is done by designing[7]:

> Design is a human activity in which we create plans[8] for the creation of artefacts that aim to have value for a prospective user of the artifact, to assist the user in his/her effort to attain certain goals. (Dorst and Van Overveld, 456)

How engineers succeed at creating this world that never was, is a rising topic in the philosophy of science, witness part III in the Handbook of the Philosophy of Science Volume 9: Philosophy of Technology and Engineering Sciences. In attempts to understand design practice, philosophers have studied whether the activity is best characterised as problem-solving (Buchanan), whether it is rational (Kroes, Franssen and Buciarelli), how to reconcile different goals (de Vries),.. I cannot go into these questions here. What is important for my point, is that several of these philosophers have drawn focus to design as a primarily synthetic nature. In his introduction to part III of (insert ref), Peter Kroes stresses the synthetic side of designing:

> […]engineers need to have synthetic design skills: when designing new technical artefacts, they must be able to combine elements (components or processes) in inventive, creative ways so that they can satisfy practical means-end or functional requirements. The designing of technical artifacts is considered to be primarily a synthetic rather than an analytic activity. […] For these purposes they also need to have synthetic skills; theories, experiments as well as experimental equipment are composed of different elements (like, for instance, laws, actions and physical components) and they result from

extrapolation, yet some authors claim it is distinctive (Illari & Russo 2014, 18). However, this does not matter for the current point.

[5] Naturally, these already existing things can *evolve*, which gives rise to change. Yet this is not the focus of the matter. This evolution is outside our control. I whish to focus on our *creation* of new things.

[6] This quote focuses on the distinction between science and engineering. Mieke Boon has on several occasions (2011a, 2011b,…) stressed that we also need to acknowledge engineering science as a scientific practice. She contrasts this with engineering practice, which is arguably also the practice Von Karman had in mind.

[7] In the literature, there is no consensus on the definition of design (see e.g. Buchanan 2009). Yet Dorst and Van Overveld give us a comprehensive one that is sufficient for the current purposes.

[8] Following Kroes, I use *design* to refer to the "plan or description" (2009b, 513) of an artefact, rather than to the artefact itself.

> researchers putting these elements together in specific ways to satisfy requirements, cognitive and otherwise. (2009a, 405)[9]

According to Richard Buchanan, the characterisation of design as synthetic activity can be traced back to Herbert Simon and refers to the idea that in designing, engineers put things together to create a functioning whole:

> He designs by organizing known principles and devices into larger systems. (Buchanan, 425)

Although it is an intuitive notion, 'synthetic' can have many different meanings. Yet design is a synthetic activity[10] in a very specific sense: it

> […] involves the synthesis of functional components that together realize the overall function of a technical artifact. (Kroes 2009a, 406)

According to Kroes, this putting together of components to realize an overall function of an artefact, is something that makes the designing of technical artefacts "a synthetic activity with distinctive features of its own" (ibid.). I agree with this conclusion.

The scientific questions that form the core of literature on induction and extrapolation, do not straightforwardly have this distinctive synthetic side. This results in differences: because we synthesize artefacts, we know the designs – we built them – and therefore have more control. Because of their artificial nature, we are faced with less ethical concerns than when we are dealing with organisms. Literature on induction and extrapolation mainly surround scientific questions in fields where we hit more cognitive (we do not have as much knowledge of the organisation of an organism as we have of a human-designed artefact) and ethical limitations (witness the ethical discussions on genetic manipulation or genetic choice). So when focusing on artefacts and designs, new questions can rise. One of them is how knowledge from known objects guides us in synthesising new ones. Though the traditional problems of induction and extrapolation hold in failure analysis as well, I want to focus on design; on how we create something new using knowledge of current things. This is what I will refer to as the *design perspective*.

2.3 Failure analysis as generalisation problems

Failure analysts proceed from causal claims regarding a specific failed artefact to causal claims regarding other types of artefacts and artefacts-to-be-developed. For their investigation, they benefit from the control and knowledge provided by designs of the artefact that failed and arrays of lab tests. The resulting causal claims have the form of recommendations and are aimed at altering the processes of use of existing artefacts, or designing new, non-existing artefacts. Unfortunately, analysts are often unspecific regarding the domain of their claims - viz. for which artefacts their recommendations hold. Yet, the objects that form the domain of their conclusion, determine what evidence the analysts need to put forward to warrant their conclusion. Whether they know the designs of the intended artefacts, whether they know the context they will be placed in etc. determines what evidence is required. For example, if analysts use one artefact failure to formulate

[9] This does not mean that designing is solely a synthetic activity. Yet the synthetic aspect is what I focus on here.

conclusions or recommendations regarding all artefacts of a certain class (e.g. all spray driers), they will need other evidence to warrant their claims than if their conclusion only applies to other artefacts of the same type (e.g. other spray driers constructed according to the same design). Given the differences that can pertain within a certain class of artefacts (they can have different designs, other materials, different functioning, etc.), warranting a claim regarding the entire class is not an easy task. This is not merely a theoretical concern. As will become clear in section 3,we can isolate inferences from failure analysts that differ with regard to base and target and therefore require different types of evidence. Yet all the recommendations give us guidance regarding what to do, what to change. Taking the design perspective, I focus on what recommendations tell us regarding how to combine specific components to create a larger whole with an envisioned[11] function. Given that one of the analyst's aims is to specify design recommendations, failure cases prove significantly insightful to study the way in which current knowledge guides the design of new artefacts.

3. Three examples of failure analysis as knowledge generalisation

In this section, I present three examples of failure analysis and flesh out three distinct types of inferences. This will help illustrate what studying these generalisations can teach us.

Example 1

Talesnick and Baker in "Failure of a flexible pipe with a concrete liner" (2001) present an analysis of a steel sewage pipe with a concrete liner, buried in a clay soil profile, located in Israel. The pipeline never got used because of severe cracking of the inner concrete liner. In their paper the authors want to

> […] determine the cause(s) of damage and the areas responsible. […]. (2001, 33)

Talesnick and Baker describe two types of tests: laboratory tests and field tests. In the laboratory test conducted on parts of the pipe, they determined the stiffness, and the vertical deflection or strain, which "induces cracking in the inner pipe line and collapse loads" (ibid, 34). It was found that

> Severe cracking of the inner liner wall (defined as a crack opening of 0.3 mm […]) occurred at a vertical diametric strain of approximately 1.2%. (ibid)

This was compared with the measurements made in the field:

> The vast majority of field measured pipe deflections […] exceed the 1.2% limit found to induce severe liner cracking of pipe sections in the laboratory. As a result the extensive damage observed in the internal pipe liner in the field […] is not surprising. (ibid, 37)

They furthermore argue that

1. Most steel pipes are considered to be flexible and designed accordingly
2. A pragmatic literature based criterion for flexible pipes is a pipe that can withstand a vertical deflection of 2% without damage

[11] Dorst and Van Overveld frame design as an intentional activity (467). I feel this is a very insightful way to understand design and which sidesteps some issues regarding goal-directedness. See their paper for more detailed information.

3. Though the pipe in question was able to withstand this, the inner liner was not, since it showed cracks at a lower vertical deflection. (ibid, 37)

One of their conclusions reads:

> ''Flexible" pipes with rigid liners must be designed with care. Flexible pipe design methodologies may be applicable, [provided that] […], the deformation limitations of the liner [are][…] carefully considered. (ibid, 42- 43)

This is how the reasoning process was laid out in the paper. Let me first attempt to present it in a logical sequence that draws focus to the different research stages. They say that most steel pipes are considered to be flexible. This entails that they should be able to withstand a vertical strain of 2%[12]. This seems based on engineering knowledge of the authors regarding the properties of steel pipes. They also state that the pipe in question is a borderline case, since it failed under circumstances that would not cause damage to a flexible pipe (2001, 33). So their reasoning can be presented as follows:

1. We assume that flexible pipes have characteristics such that they do not experience damage (viz. retain functional and structural integrity) from strain less than 2%. [assumption]
2. Bending tests on pipe segments in the lab show that the inner liner cracked at a vertical strain of 1,2%. Higher deformation can cause cracking. [tests in the lab]
3. We measured deformation of more than 1,2% in pipe segments in the field. This deformation is significant when analysing the cracking. [measurement in the field]
4. The cracking happened with deformation within the norm for flexible pipes. So even if the pipe itself was correctly designed according to flexible pipe criteria, the inner liner did not perform adequately under the specific circumstances. [inference]
5. If we want to use flexible pipe design methodologies in pipes with inner liners, we have to take the strain limitations of the liner into consideration (see quote above). [recommendation/conclusion]

I represent the base (the artefact which was the subject of the failure analysis) and target (the artefacts they mention in their recommendations) of the inference:

Base: one pipe
Target: flexible pipes with rigid liners

Example 2

In "Creep[13] failure of a spray drier", Carter (2001) presents a failure analysis of a spray drier at the Western Platinum Mine, in Rustenburg, South Africa. A spray drier is an artefact which "dries a finely divided droplet by direct contact with the drying medium (usually air)" in a short retention time (Konsidine & Kulik, 5130). The failed spray drier consisted of a cylindrical shell, with an annular gas chamber encircling the base of the shell. Four columns supported the shell. The spray drier suddenly collapsed after 20 years of service while operating normally[14] (2001, 73). The aim of Carter's investigation

[12] For more information regarding stress and strain, see the appendix.
[13] See the appendix for information regarding creep damage.
[14] It is not clear what the author means with "operating normally". Arguably, this is a judgment based on his background engineering knowledge.

> […] was to explain the failure and to make recommendations to ensure that it was not repeated on the two remaining driers […]. (ibid)

The investigation found no significant corrosion, the material was found to be accurately chosen without deterioration (ibid). Neither was there evidence of fatigue, fracture or creep damage. However, there was

> […] clear evidence of a localised buckling deformation in columns and shells in the region of the welded column-shell joint. (ibid, 74)

Carter's failure analysis methodology consists of comparing "stresses at critical points in the structure with allowable and failure stresses" (ibid, 75). Carter inferred the allowable and failure stresses from the design code for pressure vessels. He determines the influence of creep conditions on the maximum stress in the structure, both of the column-shell connection and the gas duct. Carter specifies that these calculations are estimates, yet that they "clearly indicate the nature of the failure" (ibid, 77). He calculates that the maximum stress of the structure under creep conditions is significantly above the allowable stress and the failure stress. Based on these findings, he concludes that

> The collapse of the spray drier after 20 years in service is an unusual example of a low stress, high temperature compression creep failure. (ibid,77)

Carter measured temperatures in the structure around 300°C , but argues that these are not correct by referring to the design of the spray drier. So Carter argued that the estimated temperature should be considered above 480°C. Even though creep actually redistributes stresses (ibid, 76) and thus decreases the maximum stress on the structure (compared to the stress values under elastic conditions at 500°C), the resulting stresses were still significantly above the failure stress. According to Carter, this explains the absence of clear creep rupture, and the collapse of the spray drier.
Summarizing the reasoning that led him to this conclusion:

1. He assumes that the temperatures inside the shell are higher than measured, based on the working and design of the spray drier. So creep conditions might apply, contrary to what was expected based on the measurements. [assumption]
2. Creep distributes stresses, so that the actual stress is less than the calculated elastic stress. For a high creep exponent, the maximum value of stress is about 67% of the maximum elastic stress. [background engineering knowledge]
3. In the gas duct, the stress concentration factor under elastic circumstances is 8.9, so this will be redistributed by creep circumstances to 67% of 8.9: 5,9 or, rounding up, 6. [measurement + calculation]
4. The stress concentration factor of 6 in creep conditions generates a stress value of 22 MPa. The allowed stress value is 13 MPa, failure arises at 17 MPa. [calculation]
5. Since the generated stress value was significantly above the allowed values, this resulted in fractures and collapse of the spray drier. [inference/conclusion]

Furthermore, he makes the recommendation of removing the "lagging and cladding in the region of the annular gas duct and the column-shell joints", in order to "avoid a similar fate on other more recent (and stronger) spray driers" (ibid, 77). This refers to the drier, which was "lagged and clad from top to bottom to conserve energy" (ibid, 73). The established isolation of the drier provides

evidence for his hypothesis regarding the temperature measurements. With this much isolation, he argues that the temperature was probably higher than value he measured. Similar to example 1, Carter does not regard the failure of this spray drier as an exception: he makes recommendations regarding other spray driers. He considers his knowledge about what led to the collapse of the analysed spray drier applicable to other spray driers. Yet Carter goes even further in his conclusions. He argues that the purported causal claim also holds for "more recent and stronger" spray driers. The finding that the other spray driers were also lagged and cladded from top to bottom, functions as evidence for this conclusion. The inference I want to draw attention to can be represented in the following way:

Base: 1 spray drier

Target: 2 newer and stronger spray driers with lagging in the same context

Example 3

James describes the failure of a raise boring machine in his article entitled "Catastrophic failure of a raise boring machine during underground reaming operations" (2001). Raise boring is a technique found in underground mining operations, used to produce "interconnecting vertical […] channels (raises) between underground levels in mines" (ibid, 159). The process can be characterised by two operations: drilling a pilot hole and back reaming:

> During the pilot hole drilling cycle, drill rods connect the raise boring machine with a bottom-hole assembly consisting of ribbed stabilizers, roller reamer and pilot bit. […] After the pilot hole has been completed, a raise boring head is used to back ream the required raise between the underground levels. (ibid)

Back reaming is a technique used to "increase the diameter from that initially drilled" (Slaughter e.a., 12). So the raise boring machine first drills a smaller channel connecting the two underground levels, after which a reaming head is placed on the machine (on the lower level), which is then pulled back up. The reaming head has a broader diameter than the initial drilled channel and rock cutting abilities, so the pulling up of this head creates a hole of increased diameter compared to the original one.

In his article, James describes the failure of such a machine after 119m of reaming, due to the breaking of all 32 bolts on the raise borer drive (ibid,160). He describes the site visit, inspection of the fractured bolts and the metallurgic examination consisting of chemical analysis, scanning electron microscopy, optical microscopy and hardness testing. Based on these investigations, he argues that:

> (1) The catastrophic failure of the raise boring machine is associated with the fracture of the 32 drive head bolts. Thirty of the bolts have failed as a result of corrosion-induced fatigue.
>
> (2) The bolts have failed due to a combination of high cyclic stressing induced by the operation of the equipment at 13% above maximum thrust and corrosion from the water in the flushing system. (ibid, 168)

The two other bolts had failed by 100% tensile overload (ibid, 163). He furthermore makes several recommendations:

> (1) To prevent corrosion of the bolts the following measures are recommended:

(a) An oil-based red lead primer should be used to create a barrier at the cover-body connection.
(b) Mains water should be used at all times for flushing.
(c) Equipment should not be stored underground for any length of time.

(2) Excessive thrust pressures during operation should be avoided, i.e. the equipment should be used within the limits for which it was designed. (ibid, 168)

Contrary to the example above, there is no mention of the specific artefacts these recommendations apply for. His claims appear to include more than a machine nearby. To assess his recommendations further, we need other findings which he mentions throughout the paper:

1. The "centre-bolt" torque was found to be well below the normal figure during dismantling. This could have had the effect of allowing more vertical movement of the drive head cover. (Ibid, 166)
2. Since the bolts show signs of pitting corrosion, the anti-seize compound with which they are coated "does not afford protection to the surface of the bolts". (ibid, 167)

I summarize his reasoning:

1. Tests show that the failure was due to fatigue. [tests + background engineering knowledge]
2. All fatigue areas of the 30 bolts showed signs of pitting – associated with corrosion. [tests]
3. The anti-seize compound with which the bolts were coated thus clearly did not prevent corrosion. [inference]
4. The thrust during operating was 13% above maximum, putting greater stress on the weakened bolts. [measurement + inference]

The inference can be represented in this way:

Base: one failure
Target: all (future) bolts in all possible circumstances

Summarizing the three inferences I fleshed out:

(1) this artefact (flexible pipe with rigid liner) to other artefacts of the same type
(2) this artefact (spray drier) to other artefacts with known differences (newer and stronger)
(3) this artefact (raise boring machine) to other artefacts not yet in existence

All of these inferences are related to induction and extrapolation (and external validity), but as I mentioned, they differ in one important aspect: they prescribe, among others, alterations to the design and use of artefacts. More specifically, the *first* looks like induction specified above, because base and target are of the same type. Yet the result of the inductive step is a design recommendation (how to choose the liner), instead of a general claim about the entire pipe. This is a significant difference with induction as specified above. So this looks like induction, but the focus on redesign differentiates it from the generalisations discussed in section 2. The result is a prescribed action. The *second* inference is better characterised as a sort of extrapolation, because of the known differences between base and target. Yet the conclusion is again focused on redesign of the target, something that is not represented in the literature on extrapolation. Note that the differences concern properties of the target. They can also take the form of usage or context. In this case, the context is explicitly stable: the two newer spray driers are located on the same site as the failed one. This is an

important part of information contained in the inference and needs to be represented in our analysis. A similar point holds for similar maintenance practices. The *third* inference is similar to the second, only this time, there is no context specified in the recommendations. The inference apparently does not depend on the context the artefacts are placed in. Moreover, it does not merely apply to artefacts that are at the time of the analysis in the vicinity of the analysed artefact, but also to artefacts-to-be-designed. This might seem significantly different from the other examples, but is arguably related.

These inferences are thus arguably slightly different generalisations and all involve design recommendations. In the remainder of this paper, I will attempt to get a more profound understanding of whether, why and when they are warranted. These are important questions, since implementing changes in designs is not that straightforward:

> The complexity of many engineered artefacts, together with their interactions with a changing environment, make working out the effects of many design changes either analytically intractable or analytically very difficult [Pavitt, 1984; Nightingale, 2004]. (Nightingale, 365)[15]

The influence of rather 'small' changes, like whether centre bolt is torqued to the normal figure (see example 3) can be enormous. Putting together different components in such a way that they combine to realize the envisioned behaviour, is a complex and open-ended process (Dorst and Van Overveld, 456). Making sure that an artefact functions as envisioned therefore often involves "learning, experimentation, testing, and numerous modification and feed-back loops" (Nightingale, 365). One of the learning occasions is the failure of artefacts. Understanding how knowledge from other, failed artefacts can be implemented in the fickle balance that is created in designing new artefacts, will require different tools than understanding generalisation in scientific contexts that do not focus on this *design-perspective.* Note that the examples also show that the balance is indeed more fickle than suggested by Cartwright and Hardie (see the introduction). Not only can small changes have big consequences, artefacts can also fail after years of functioning normally, like the spray drier in example 2. So it can appear that everything is present for correct and stable functioning, yet after time, in certain contexts, it turns out that something was missed. Knowing how and why artefact failures occurred, when they can occur again and how to repair artefacts and alter designs to prevent failures, is therefore also not a straightforward matter.

4. Analysis of the examples

So these recommendations are not unproblematic. More specifically, they only hold if certain stable causal relations hold. For example, the recommendation

> An oil-based red lead primer should be used to create a barrier at the cover-body connection for all artefacts of type X.

only holds if there is some kind of stability in causal factors across different artefacts and contexts:

[15] In that sense, it is not surprising that the design perspective has not been excessively studied in the generalisation literature. Many philosophers studying the generalisation problems mentioned in section 2, focus on the biomedical and social sciences. Because of ethical limitations and cognitive constraints I mentioned already, successfully changing the functioning of an organism is even more difficult than changing the functioning of an artefact.

> For all artefacts of type X, an oil-based red lead primer is a negative causal factor for bolt breakage.

Intuitively, this comes down to saying that oil-based red lead primers can (and sometimes do) prevent bolt breakage. The term 'causal factor' will be defined more thoroughly below. For now, it is important to see that these causal factors need to be stable in some sense, if they can ever be the base for generalisations. Otherwise, we can never safely assume that they will hold in other situations. This is crucial to understand my case studies, since formulating recommendations from one failure is, as mentioned in the introduction, a type of knowledge generalization. From the overview in section 2, it is clear that knowledge generalization is not without problems. This is also the case in failure analysis. The analysts need to provide reasons *why* their generalisation is warranted. They also need to provide arguments for the generality of their conclusion: the artefacts that they consider part of the domain of the claim determine its validity. To explain this, I adapt example 3. Suppose that James based his analysis on multiple failures of raise boring machines that have the same design, call this type T_1. James then formulates the same recommendation as in the original article, namely

> (a) An oil-based red lead primer should be used to create a barrier at the cover-body connection.

Suppose that there are raise boring machines with stainless steel bolts (type T_2). These bolts are not susceptible to corrosion in the same contexts as other bolts, so the causal claim does not hold for these machines in similar contexts and correspondingly, the recommendation would be irrelevant. This also shows how we can intuitively understand the stability mentioned above: as related to capacities of (parts of) artefacts in certain contexts. The stainless steel bolts (in the same context) do not have the increased capacity to corrode. The pipe in example 1 has an increased capacity to bend, while not all other pipes do. I will present elaborate and theoretical underpinnings of capacities in section 5. Here, I want to reflect on the need for justification for generalizations, and the required reference to the implied domain.

Steel's discussion (2007) of extrapolation in the biomedical sciences can deepen our understanding of these challenges. According to him, a theory of extrapolation has to solve two basic challenges: the extrapolator's circle and the problem of causally relevant differences between model and target population (2007, 4). The first

> […]arises from the fact that extrapolation is worthwhile only when there are important limitations on what one can learn about the target by studying it directly. The challenge, then, is to explain how the suitability of the model as a basis for extrapolation can be established given only limited, partial information about the target. (ibid.)

The problem of causally relevant differences, on the other hand,

> […] is a direct consequence of the heterogeneity of populations studied in biology and social science. Because of this heterogeneity, it is inevitable that there will be causally relevant differences between the model and the target population. Thus, an adequate account of extrapolation must explain how it can be possible to extrapolate from model to target even when some causally relevant differences are present. (ibid.)

Based on the discussion of my cases, I argue that both challenges are also present in failure analysis. Analysts need to provide reasons why (1) base and target are similar and (2) account for relevant differences between base and target. Though Steel focuses on heterogeneity in the social and biomedical sciences, causally relevant differences between artefacts are as problematic – witness the

spray drier with stainless steel bolts. So it is important to specify the domain of a causal claim in order to evaluate it. It is widely recognised that we cannot safely make causal claims in the biomedical sciences without referring to populations. Yet it is equally important to make explicit what causal claims refer to in the physical sciences, as it is in the biomedical ones. Especially when making design recommendations, see sections 2 and 3. To argue that one failure is relevant for other (instances of the same type of) artefacts, analysts thus need to provide evidence. Unfortunately, none of these articles straightforwardly do this. Yet, the recommendations are successful (or so the articles mention). It therefore seems that the inferences are warranted, but the justification is not made explicit or is not reflected upon. Bucciarelli mentions a related observation:

> Epistemological questions about the source and status of engineering knowledge rarely draw [the engineers'] attention.[…] If their productions function in accord with their designs, they consider their knowledge justified and true. (2003, 1)

In the following sections, I will attempt to make the analysts' reasoning and presuppositions explicit. This will allow us to reflect more profoundly on the nature of evidence their generalisations need. For this, I will use Steel's framework (2007) as a starting point. The first step towards this reflection is reframing the recommendations as causal claims.

5. Philosophical tools for investigation: making things explicit

5.1 Capacities, features and MOD

To reframe the recommendations as causal claims, I first need a more precise definition of the notions 'causal factor' mentioned above. It is inspired by Giere, but I will connect it to Cartwright's notion of capacities. As noted in section 4, causal factors need to be stable in some sense to allow for generalisation. I represent this stability via capacities:

> All bolts have the capacity to break in some contexts.
> Corroded bolts have an increased capacity to break.
> Therefore corrosion is a causal factor in bolt breaking.

Pointing to causal factors gives engineers a way to indicate capacities, which in turn allow for generalisation. Cartwright illustrates what she means by 'capacity' via the claim that aspirins relieve headaches. According to her, this claim

> […] says that aspirins have the capacity to relieve headaches, a relatively enduring and stable capacity that they carry with them from situation to situation; a capacity which may if circumstances are right reveal itself by producing a regularity, but which is just as surely seen in one good single case. The best sign that aspirins can relieve headaches is that on occasion some of them do. (1994, 3)

I use Cartwright's notion of capacities for several reasons. One, this notion is inherently connected to probabilistic causal relations (as is causal factor).

> The point is that, for each capacity the cause may have, there is a population in which this capacity will be revealed through the probabilities. (1994, 121)

Second, capacities are context-dependant; local (Illari & Williamson, 153). Whether and how a capacity actually manifests, depends on the situation.

> […] capacities […] can be assembled and reassembled in different nomological machines, unending in their variety, to give rise to different laws. (Cartwright 1999,52)

Third, the notion of capacity is central to Cartwright's concept of nomological machine:

> […] a fixed (enough) arrangement of components, or factors, with stable (enough) capacities that in the right sort of stable (enough) environment will, with repeated operation, give rise to the kind of regular behaviour that we represent in our scientific laws. (1999, 50).

On Cartwright's account, nomological machines produce regular behaviour and correspondingly, laws (of nature) only hold "relative to the successful repeated operation of a nomological machine" (1999, 50). Arguably, all technical artefacts are nomological machines in the sense that they give rise to regular behaviour. Using Cartwright's framework of capacities thus allows for a probabilistic, local notion of causal connection. This is an elegant and useful way to reformulate the recommendations from the failure analysis, though I will make some adaptations.

One adaptation that I need is to reframe them such that the domain of the causal claims is clearly represented (cf. supra). For the current purposes, we can do this by referring to types of artefacts.

> For all artefacts of type X, c is a positive/negative causal factor for e.

Where c and e express a specific value of a relevant variable feature of (part of) the artefact, viz. the value that is thought to be important for the causal claim. I will get back to parts of artefacts later.

Consider example 1. Since Talesnik and Baker analyse only one artefact, their causal claim can be represented as follows:

(E1) For this pipe, deflection is a positive causal factor for cracking of the liner.

To allow generalisation, (E1) needs to correspond to:

(E1*) The deflected pipe has increased capacity for cracking the liner, and it probably manifested for the studied pipe system.

According to Talesnik and Baker, the capacity manifested because the liner cracked and this was at least partly due to the deflection of the pipe. These claims are token level causal claims, referring to one specific artefact. The main causal claims from examples 2 and 3 (and the corresponding capacity claims) can be represented in a similar way.

(E2) For the Rustenburg spray drier, the lagging and cladding of the annular gas duct is a positive causal factor for collapse of the spray drier.

(E2*) A lagged and cladded annular gas duct has an increased capacity for collapse of the spray drier and it probably manifested for the Rustenburg spray drier.

(E3) For this raise boring machine, the corrosiveness of the flushing liquid is a positive causal factor for bolt.

(E3*) Corrosive flushing liquid has increased capacity for corrosion, which increases the capacity for breaking the bolts and it probably manifested for this raise boring machine.

These are the claims the failure analysts implicitly start from: claims regarding the artefact they investigated. Their recommendations can be represented by referring to types of artefacts:

(E1') Deflection of a pipe is a positive causal factor for cracking of the liner for all artefacts of type X.

(E2') The lagging and cladding of a gas duct is a positive causal factor for collapse of the spray drier for all(?) artefacts of type Y.

(E3') The corrosiveness of flushing liquid is a positive causal factor for breaking of the bolts for all artefacts of type Z.

Analysts go from evidence to causal factor claims. Doing so assumes that by identifying these factors they succeed in identifying some capacity that is stable under certain circumstances. Clearly, not all pipes crack liners. In order to adequately apply the recommendations (viz. everywhere they might be useful, not where they are thought not to be) we need a clear representation of the circumstances under which the capacity can actually manifest. Cartwright's notion of nomological machine is, as such, not very helpful here. I agree that whether capacities manifest depends on the contexts and the specific machine they are embedded in, which is a big step towards the design perspective. Cartwright's work has undoubtedly been incredibly important in studying design. But from Cartwright's definition, it is not clear how we can discover which environment and arrangement of components is necessary for specific capacities to manifest in a certain way.[16] Her account does not fully embrace the design perspective described in section 2: the actual scientific practice of synthesising components into a functioning whole. So I need to specify how we can discover what the 'right sort of stable environment' and 'arrangement of components' are when designing new artefacts. Based on the examples, I argue that this requires specifying

1. the type of artefact
2. the relevant causal factor
3. the context

With 'relevant causal factor', I mean the causal factor which corresponds to the increased or decreased capacity. The first two are already present in the current formulation, but the context is not yet represented. Yet, this is an important piece of information. In (re)design literature, this is reflected in the distinction between mode of deployment (MOD) and mechanistic organisation of the artefact (Chandrasekaran & Josephson 2000). The former represents the ways in which the artefact is used, the latter the way the artefact is constructed. For the task at hand, MOD can be understood in a broader sense and also represent certain important aspects of the *environment* of use (e.g. underground, in high humidity) instead of merely *mode* of use (e.g. operating at high thrust):

For all artefacts of type X and MOD Y, c is a positive/negative causal factor for e.

The way the analyst formulate their recommendations in example 2, seems to imply that they only hold for artefacts in the same context (warm climate). In example 3, there is mention of *operating at high thrusts*, but not of requirements for context. MOD can capture both.

5.2 Failure mechanisms

[16] Cartwright does address this question in relation to social policies in her and Hardies' 2012 book. But because of the specific synthetic nature of designing and the greater need for information, their framework is not adequate for the current purposes. They moreover do not take the design perspective. So in order to understand how failures of existing artefacts shape the design of new ones, I cannot use their framework.

So it is clear that failure analysts have certain (un)specified beliefs regarding what factors are relevant to warrant the causal claim. The tools I have presented help to make them explicit. We can now turn to the question of whether and when they are justified. This comes down to determining how to characterize "type X" in the definition above. For this, I will present a two-fold mechanism-based procedure.

Representing failure analysis in terms of failure mechanisms allows me to provide a fruitful answer to the challenges raised in section 4: the extrapolator's circle and relevant differences relating to the design perspective. Moreover, a mechanism-based framework fits well with my characterisation of causal claims in terms of capacities. Cartwright recently connected her notion of nomological machines to the mechanism literature (2009, 7). According to her, we can understand mechanisms as nomological machines. In this way, her work on nomological machines functions as a connective between capacities (that allow me to express required stability demand) and mechanisms (that can form the basis of the generalization procedure). Furthermore, the term "mechanism" is often used by failure analysts themselves. For the characterisation, I borrow the general definition of a mechanism from Illari and Williamson (2012):

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (123)

The examples above all constitute a mechanism in this sense: they refer to entities (e.g. the pipe) and activities (corrode, break), organised in a specific way (the liner covers the inside of the pipe, the bolts hold the driver head in place) such that they are responsible for a specific phenomenon (the failure of the pipe, the collapse of the drier). I find the definition of activities as "producers of change" by Machamer, Darden & Craver (2000, 3) most appropriate to capture the activities at hand. They are "usually designated by a verb or verb form" and "are constitutive of the transformations that yield new states of affairs or new products" (ibid, 4). Clearly, corroding and breaking satisfy this definition. MDC furthermore talk about "bottoming-out":

> Different types of entities and activities are where a given field stops when constructing mechanisms. The explanation comes to an end, and description of lower-level mechanisms would be irrelevant to their interests. (ibid, 13)

This is also the case for the verbs I identified as activities in failure analysis: the mechanism of generating stress or corroding is not spelled out in the analyses. These activities arguably constitute the bottom level in failure analysis. The same holds for the entities involved in their reasoning: analysts do not provide explanations in terms of e.g. atoms or molecules. Finally, because Steel (2007) formulates a mechanism-based strategy to the challenges raised in section 4, I can use his work as a starting point. Characterizing failure analysis in terms of failure mechanisms is thus a promising choice. In the next section, I present Steel's framework before adapting it to fit the failure analysis examples.

5.3 Steel on comparative process tracing

The strategy Steel develops is called comparative process tracing (CPT):

> First, learn the mechanism in the model organism, by means of process tracing or other experimental means. Second, compare stages of the mechanism in the model organism with that of the target organism in which the two are most likely to differ significantly. (ibid, 89)

Steel distinguishes two steps to CPT. The first (process tracing or other experimental means) deals with mechanism discovery.[17] In failure analysis this often has to do with background engineering knowledge, such as the properties of flexible pipes. I will not discuss this further.

The second step is relevant for the current purposes: look for significant differences between model and target. If significant differences pertain, we cannot be sure that the behaviour of the model will be replicated in the behaviour of the target:

> Significant differences are those that would make a difference to whether the causal generalization to be extrapolated is true in the target. (Steel 2007, p. 89)

To check for such differences, we need "generalizations asserting that objects of specified types resemble one another in certain ways though not necessarily in others" (ibid.). Knowledge of these generalizations allows us to (1) check in a directed way whether the specific differences occur and correspondingly (2) judge whether the extrapolation is warranted or not. Steel developed his framework for the biomedical and social sciences, so he is looking for generalizations like

> Features A,B, and C of carcinogenic mechanisms in rodents usually resemble those in humans, while features X,Y, and Z often differ significantly. (ibid)

Because of my focus on to-be-designed artefacts and design recommendations, Steel's generalisations will not do the job. The comparisons Steel suggests are apt to capture biomedical examples, but are too unspecific to reflect the amount of control we have over, and knowledge we possess of, artefacts. Since we have more knowledge of and control over artefacts, we *can* compare existing to non-existing artefacts in a more specified and detailed way than Steel allows in his framework: we can compare designs on specific points for example. But more importantly, because of the difficulties of adapting designs (see section 3), we *need* to be specific and warrant the applicability of recommendations thoroughly. Uninformed or unspecific implementation of recommendations can have unforeseen consequences; small changes can result in grave problems. Moreover, though organisms evolve too, Steel's account is not focused on the required knowledge to *actively change* organisms[18] – he does not take the design perspective. Yet this is exactly what I am interested in regarding artefacts. So like with Cartwright, Steel's CPT provides a good basis to model generalisations to non-existing artefacts, but needs some changes and additions.

In the following sections I show how Steel's framework can be adapted to fit failure analysis. I will first elaborate on what it means for artefacts to be similar, and present a mechanism-based account for that. I then proceed to adapt and apply Steel's CPT to fit the failure analysis examples.

6. A mechanism-based generalization framework

6.1 Similarity

[17] Process tracing refers to two strategies for discovering mechanisms: schema instantiation and forward chaining/backtracking.

[18] There is the question of genetic modification which I mentioned in section 2. I am not focused on this scientific practice, but I believe my account (and the adapted version of CPT that I present) has the potential to be useful in this context as well. I get back to this in section 7.

CPT depends on knowledge of likely similarities and dissimilarities between base (e.g. mice) and target (e.g. humans). But before the actual CPT can begin, we need to ensure that there is enough similarity between base and target to allow the base to function as a suitable model for the target. In biomedical sciences, this comes down to knowledge of some mechanism e.g. the metabolism in mice and humans. I argue that, in failure analysis, this comes down to knowledge of whether the failed submechanism of the artefact is present. It refers to a submechanism which helps sustain the artefact and its functioning. In the example of the raise boring machine, the submechanism (M) that failed is the mechanism attaching the cover of the boring head to the body. It can be characterised as follows:

> Entities: connecting parts, cover, body
>
> Activities: connecting parts immobilize cover, cover is immobilized
>
> Organisation: cover is fastened onto body via connecting parts

Every artefact containing a submechanism of this type is a candidate for the domain of the recommendation. Note that this implies that different mechanisms containing submechanisms of the same type can be part of the domain. Remembering the characterisation of causal claims underlying recommendations I discussed in the previous section, the relevant question we need to ask is:

> For all artefacts containing a submechanism of type M and operating in MOD Y, is c a positive/ negative causal factor for e?

In the following section, I answer this question by adapting CPT.

6.2. CFPT – Comparative failure process tracing

I agree with Russo and Williamson (2007) that the

> […] existence of a mechanism provides evidence of the stability of a causal relationship. […] In other words, mechanisms allow us to generalise a causal relation. (159)

Cartwright makes a similar claim with regard to nomological machines:

> […] laws of nature obtain […] on account of the repeated operation of a system of components with stable capacities in particularly fortunate circumstances. […] it takes what I call a *nomological machine* to get a law of nature. (1999, 49)

This is also the case in the generalisations I am concerned with. Adapting Steel's CPT to fit the generalisations from my examples will shed light on how existence of a mechanism provides evidence. Recall that Steel (2007) developed his framework mainly to deal with organisms. Compared to organisms, there is a crucial difference to analysis of artefacts: we know the designs – we built them. This allows for greater manipulability. Where Steel, for the biomedical examples, has to refer to "knowledge of likely dissimilarities", we can be more specific with regards to the nature of these dissimilarities in failure analysis. I argue that for the failure mechanism[19], we need 3

[19] The similarity criterion referred to a submechanism partly responsible for artefact functioning. The failure mechanisms is responsible for the failure phenomenon. These are two different phenomena, and thus call for different mechanisms.

comparison points and a check for counteracting mechanisms. I will first discuss the 3 comparison points. They are

(1) types of parts
(2) organisation
(3) activities and corresponding properties

The connection between activities and properties fits my definition of activities in the MDC sense:

> […] activities determine what types of entities (and what properties of those entities) are capable of being the basis for […] acts. Put another way, entities having certain kinds of properties are necessary for the possibility of acting in certain specific ways, and certain kinds of activities are only possible when there are entities having certain kinds of properties.. (2000, 6)

It furthermore ties in with the capacities that ensure the stability of the causal factors. So this is a fruitful connection. Call the combination of our 3 comparison points and counteractive mechanism checking 'Comparative Failure Process Tracing' (CFPT). I now illustrate these comparison points via the examples in order to facilitate arguing for each of them.

Example 1: The pipe system.

This is an interaction between instances of two types of parts: a pipe and a liner. They are organised in a specific way: the liner covers the inside of the pipe. They furthermore interact such that the pipe deflects and the liner responds by cracking. This interaction is connected to specific properties of the pipe and the liner: the pipe is flexible, the liner is rigid.[20] There is no MOD specified – Talesnick and Baker argue that the pipe-liner interaction is due to a design problem. Representing these aspects in a standard format:

Types of parts involved: pipe, liner
Properties: pipe is flexible, liner is rigid
Organisation: liner covers inside of pipe
MOD: no usage or context specified

Similarly:

Example 2: The spray drier

Types of parts: cladded and lagged shell of a spray drier, gas duct
Properties: the shell has a mass, the gas duct can break
Organisation: the shell leans on the gas duct
MOD: no specified usage, creep temperatures

Example 3: The raise boring machine.

Types of parts involved: parts that immobilize cover of the drive head, flushing liquid
Properties: flushing liquid is corrosive, immobilization parts are susceptible to corrosion
Organisation: flushing liquid engulfs immobilization parts
MOD: high thrust, no specified context

[20] There are threshold values, but this does not matter here. There is no reference to threshold values in the recommendations of the failure analysts.

With this in mind, I can argue that these points can house significant differences.

1. Types of parts:
Suppose we know that a specific raise boring machine does not use cooling liquid (T3). The type of part (cooling liquid) is not represented. Therefore we cannot say that the failure mechanism will also take place in T3 raise boring machines. There is no entity to partake in the activity that is crucial to the failure mechanism (viz. corroding the bolts). In this case, there are even reasons to believe the mechanism will not be active.

2. Properties:
As MDC stated, entities partake in activities because of some properties - they need to have the capacity associated with the causal factor specified in the causal claim underlying recommendation. Even if all entities are present in the target artefact, they need to have the required properties to take part in the relevant activities. I already mentioned one example in section 4, when talking about a raise boring machine with stainless steel bolts rather than bolts that can corrode. Another example would be a spray drier with an unbendable and unbreakable gas duct. It will not be susceptible to the same failure mechanism as is present in example 2, since the gas duct cannot break. The shell of the spray drier still has the capacity to break gas ducts, but it will not manifest because the gas duct does not have the capacity to break. If the liner of the pipe is not made of concrete, but instead of some other flexible material, it will not crack.[21]

3. Organisation:
This is fairly straightforward. If the organisation of the entities differs significantly, the failure mechanism will not be active. If the shell does not rest on the gas duct, it will not generate stress on the duct and the same mechanism can therefore not be said to hold. Other mechanisms can of course be present that generate the same effect, but I would argue that they need other recommendations.

So these three points of comparison describe features where significant differences can pertain. Note that MOD also remains an important point of comparison, but is arguably distinct from the other points, since they deal with aspects of the failure mechanism. If no dissimilarities are found, the failure mechanism can be active. Combining CFPT with the information above, we arrive at the following characterisation of the recommendations' domain:

> For all artefacts that (1) contain a submechanism of type M, (2) are used in MOD Y and (3) pass the CFPT, c is a positive/negative causal factor for e.

Let me briefly discuss the required check for counteracting mechanisms. If, for example, the shell of the spray drier had ventilation holes while being lagged and cladded, the failure mechanism might also not be active. Determining what counts as a counteracting mechanism again depends on a lot of background knowledge and applications of multiple scientific regularities. To illustrate, consider a submechanism that is placed in a new artefact, but behaves completely different there; in an

[21] The phrasing of example 1 confirms this point: they make explicit reference to flexible pipes with rigid liners, implying correctly that their claim does not necessarily hold for non-flexible pipes and/or non-rigid liners.

unforeseen way[22]. Based on the discussion in section 3, this is a real possibility. This means that there are causal relations that we have not taken into account. Specific parts, properties, features of the organisation or specific counteracting mechanisms have not been considered in designing the new artefact. This 'failure' of the submechanism teaches us about new causal relations, about mechanisms that we did not consider to be counteracting, about connections that we considered negligible but weren't. Designing an artefact is attempting to make a nomological machine, a deterministic system which behaves as we want it to and not differently. If something does not behave the way we envisioned, we missed something in the description or shielding. My framework allows the engineers to specify what happened and why, instead of just acknowledging something somewhere went wrong – which is arguably important to make warranted decisions and act in warranted ways. Yet this is no plea for infallible designs. I agree with Bucciarelli that

> […]there will always be a potentially problematic state of affairs not considered, overlooked, unimagined, unconstructed, no matter how many safety procedures one invokes or how imaginative and free wheeling your brainstorming session about possible contexts of use may be. (2003, p. 30)

There can always be aspects that haven't been taking into account. So looking for a procedure to provide a definitive answer regarding whether a specific failure will occur, is a futile undertaking. I have therefore not described an *algorithm* for determining the domain of failure recommendations, but rather presented a *heuristic* to determine whether recommendations are relevant. As I mentioned already, we often have specific, reliable and direct ways to check the features mentioned in the heuristic for specific artefacts: designs. Information regarding types of parts, properties of these parts, organisation and possible counteracting mechanisms are often mentioned there. So by representing the mechanisms that failure analysts identify as responsible for the failure phenomenon in a way that highlights these comparison points, engineers can actually learn from past failures in an easy way and thoroughly check whether the recommendations are relevant for their specific situations. In biomedical sciences, CPT informs us which model population will succeed most in capturing the mechanism in the target:

> Thus, comparative process tracing yielded the conclusion that the rat was a better model than the mouse. (Steel 2007 p. 91)

The target population in biomedicine is often humans. In failure analysis, on the contrary, what the model teaches us determines the target; the domain of the analyst's recommendations.

6.3 Relation to Cartwright and Steel

Now that my framework is completely spelled out, I can further specify why Steel' and Cartwright's notions did not suffice to capture how we generalise to non-existing artefacts, and correspondingly, how we can use failure of existing artefacts to create new things. As I mentioned in section 5, Cartwright's capacities allow for a local, probabilistic notion of causality. This was very useful to characterize the causal claims from the failure analysis examples. Yet Cartwright's notion of capacities (and the related notion of nomological machines) as such cannot characterize when inferences to non-existing artefacts are warranted. Capacities and nomological machines are not specific enough for this goal. Consider the example of the aspirins from section 5.1:

[22] I thank an anonymous reviewer for suggesting this.

> The best sign that aspirins can relieve headaches is that on occasion some of them do. (Cartwright 1994, 3)

Yet when we want to design a new type of aspirin, we need to know the specific circumstances under which they relieve headaches, so that we can ensure that the newly designed aspirin will also manifest this capacity. Clearly, we need knowledge of capacities for this, but that is not all, we need more. Besides the capacities, we also need specific information of the environment and arrangement of components needed to make the capacities manifest. Only then can engineers (or chemists) attempt to successfully synthesize components into a larger system with a specific function – which is what it means to design an artefact (recall section 2). Cartwright's discussion of nomological machines touches on this (e.g. Cartwright 1999, 64), yet does not give us specific guidance as to how we should collect or present this information. My framework, on the other hand, gives insight into the nature of the knowledge required by the design perspective and facilitates its presentation via a mechanism-based procedure. It is not surprising that Cartwright's account cannot answer the questions I am concerned with. Her main goal is to refocus the debate on laws to capacities by arguing for "a patchwork" of laws, instead of a pyramid. She developed 'nomological machines' for this goal. As such, it is not sufficiently specific to guide the specific question of how we can develop new artefacts from failed ones.

A similar point holds for Steel. When focusing on how we succeed in designing new things, we need significantly more information than merely reference to likely similarities and differences in the operating mechanisms. As I argued in section 6.2? we need certain specific things to stay the same: the parts, their relevant properties, the organisation, the mode of operation. It is of utmost important to specify these comparison points if we want to understand why failure analysts can make recommendations regarding objects-to-be-designed. Above that, we need a way to specify the role background engineering knowledge plays. So all of this 'messiness' cannot be captured by referring to similar mechanisms and likely differences. In creating new things, our control is greater, but the amount of required evidence to warrant generalisations, is as well.

7. Conclusion.

I started this paper with an overview of several problems and related on-going debates regarding knowledge generalisation. Reflecting on engineering practice and specifically failure analysis, I have argued that philosophical discussions of such problems need to be expanded to cope with creation of new artefacts. In general, discourse on knowledge generalisation focuses on targets already in existence. I argued that certain reasoning (specifically relating to the designing of new artefacts) in scientific practices, including failure analysis, is not adequately characterised in this way. I argued that studying artefacts has several 'benefits': we have greater knowledge of artefacts, since we designed them. This, combined with less ethical restraints due to their artificial nature, results in greater control over them. Because of that, new questions arise. One of them, the question I focused on, asks how we can use knowledge from existing artefacts to design new ones. In other words, how can knowledge of (failed) artefacts guide us in combining functional components into a lager system with an envisioned overall function? I called this the design perspective on generalisation. Such generalisations are present in, among others, failure analysis. I have provided a first attempt to characterise these inferences and reflect on when they are justified. I illustrated this with case studies from failure analysis. I fleshed out three different types of inferences to new artefacts: one that looks like induction, one that looks like extrapolation and one that is neither. I proceeded to

analyse these inferences by representing them in a standard format based on Cartwright's notion of capacities. This allowed for probabilistic, local causal claims, while accounting for the stability required for generalisations. Because of my focus on design, I adapted Cartwright's discussions on capacities and nomological machines. In order to successfully build nomological machines (what artefacts are), we need more information than a general reference to components with stable capacities and a right sort of stable environment; we need to know what the components are and what the 'right sort' of environment is. We need qualitative information and a way to represent it. Only then can we create an artefact with the envisioned functional behaviour. I also argued that we needed to specify the mode of operation, to account for different types of use and contexts the artefact can be placed in. Combining these insights, I argued that engineers implicitly look for claims of the following format:

> For all artefacts of type X and MOD Y, c is a positive/negative causal factor for e .

I then presented a heuristic to determine which artefacts belong to 'type X' – the domain for which the inference is valid and what evidence we need for this. For this, I used and adapted Steel's mechanistic framework of warranted extrapolation. It hooked nicely onto the mechanistic representation of artefacts I presented. It depends on "likely similarities and dissimilarities of base and target". Like with Cartwright, my focus in artefacts and design demanded adapting Steel's framework. Because of the specific synthetic nature of designing and the complexity of changing designs, I argued that we need more specific information to determine when recommendations are warranted for artefacts-to-de-designed. Fortunately, we also have more knowledge of artefacts, so we can provide this information. Starting from these insights and the examples from failure analysis, I argued that we can develop a more specific description of what it means for artefacts to be similar or different in ways relevant to the inference. Regarding similarity, I argued that (new) artefacts are candidates for the domain of the inference if they contain the submechanism which failed in the original artefact. I represented this in the following way:

> For all artefacts containing a submechanism of type M and operating in MOD Y, is c a positive/ negative causal factor for e?

I then developed a mechanistic (heuristic) procedure to check for relevant differences and thus determine (non-deterministically) a justified answer to this question. As mentioned, the artificial nature of 'artefacts' allowed for greater specificity than the cases Steel deals with. I argued that to determine whether the study's failure can also manifest for a certain artefact, we need to check three points of comparison, viz. whether relevant *parts* are present, whether these parts have the appropriate *properties* and whether they are *organised* in a way that is similar. Finally, we need to check for *counteracting mechanisms*. I called comparing the three aforementioned points and checking for counteracting mechanisms "Comparative Failure Process Tracing":

> For all artefacts that (1) contain a submechanism of type M, (2) are used in MOD Y and (3) pass the CFPT, c is a positive/negative causal factor for e.

I stressed that all of these steps require a great deal of background engineering knowledge and that this procedure should therefore not be seen as an algorithm, but merely as a tool for making the inferences explicit. In this way, I hope to have provided a first attempt to reflect on generalisations that deal with artefacts not yet into existence.

The account developed in this paper is relevant for both philosophers and engineers. For philosophers, it can provide input for a theory of evidence, among others on the subject of mechanisms. It also draws attention to an under investigated aspect of knowledge generalization: when and how can we generalize in order to design new objects? The analysis I presented can possibly provide inspiration for similar inferences in other innovation contexts – such as genetic manipulation and pharmacology. If medical practitioners want to engineer new drugs or chemical compounds based on knowledge we possess today e.g., they also need strategies to determine when and whether our current knowledge provides a base to warrant new designs. Moreover, by understanding the differences between technical scientific practices and social and biomedical ones, we can gain a more profound understanding of these sciences and their relations. Engineers can also benefit from this paper. For one, it allows failure analysts to present stronger arguments for their recommendations by making the required evidence explicit. My framework can even provide ways to make the analyst's recommendations more precise. By using my framework analysts can tie their formulations more clearly to the evidence that other engineers can use to evaluate the whether the recommendations are relevant for the machine and context the engineers are interested in. This is related to the importance of packaging knowledge in a way that allows travel.[23] The same can be said about representing information from failure analysis cases in such a way that engineers can reuse them in other contexts, to avoid failure or make design adjustments. Moreover, other engineers that wish to use the knowledge from failure case studies need lots of background knowledge of a specific artefact or failure sub-discipline to evaluate whether a case study is relevant for their situation. Moreover, many sub-disciplines in (design) engineering have their own methodologies and criteria (Dorst & Van Overveld 456). The procedure specified in this paper, combined with the tools to make the level of generality explicit, can aid analysts and designers in determining whether failure scenarios are applicable to their specific cases.

[23] See Sabina Leonelli's work (e.g. 2010) for detailed discussions on the importance of packaging in the biomedical sciences.

References

Bechtel, William. 2008. *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Taylor & Francis.

Boon, Mieke. 2011a. "In Defense of Engineering Sciences: On the Epistemological Relations Between Science and Technology." *Techne* 15 (1): 49–71.

———. 2011b. "Two Styles of Reasoning in Scientific Practices: Experimental and Mathematical Traditions." *International Studies in the Philosophy of Science* 25 (3): 255–78.

Bucciarelli, Louis L. 2003. *Engineering Philosophy*. Delft University Press.

Buchanan, Richard. 2009. "Thinking About Design: An Historical Perspective." In Vol. 9, *Handbook of the Philosophy of Science: Philosophy of Technology and Engineering Sciences*, ed. Dov Gabbay, Paul Thagard, John Woods, and Anthonie WM Meijers, 409 – 453. Amsterdam: Elsevier.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago, IL, US: University of Chicago Press.

Carter, Paul. 1998. "Creep Failure of a Spray Drier." In *Failure Analysis Case Studies II*, edited by D.R.H. Jones, 73-77. Amsterdam: Pergamon, 2001. Originally published in D.R.H. Jones, *Engineering Failure Analysis* 5(2): 143-147 (Amsterdam: Pergamon, 1998).

Cartwright, Nancy. 1994. *Nature's Capacities and Their Measurement*. Oxford University Press.

———. 1999. *The Dappled World: Essays on the Perimeter of Science*. New York: Cambridge University Press.

———. 2009. "How To Do Things With Causes." *Proceedings and Addresses of the American Philosophical Association* 83 (2): 5–22.

Cartwright, Nancy, and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better*. New York: Oxford University Press.

Chandrasekaran, B., and John R. Josephson. 2014. "Function in Device Representation." *Engineering with Computers* 16 (3-4): 162–77.

Considine, Glenn D, and Peter H Kulik. 2008. *Van Nostrand's Scientific Encyclopedia*. Hoboken, N.J.: Wiley.

De Vries, Marc J. 2009. "Translating Customer Requirements into Technical Specifications." In Vol. 9, *Handbook of the Philosophy of Science: Philosophy of Technology and Engineering Sciences*, ed. Dov Gabbay, Paul Thagard, John Woods, and Anthonie WM Meijers, 489 – 512. Amsterdam: Elsevier.

Dorst, Kees, and Kees van Overveld. 2009. "Typologies of Design Practice." In Vol. 9, *Handbook of the Philosophy of Science: Philosophy of Technology and Engineering Sciences*, ed. Dov Gabbay, Paul Thagard, John Woods, and Anthonie WM Meijers, 455 – 487. Amsterdam: Elsevier.

Elsevier B.V. 2016. "Engineering Failure Analysis." Accessed July 8 2016. http://www.journals.elsevier.com/engineering-failure-analysis/

Goodman, Nelson. 1955. *Fact, Fiction, and Forecast.* Cambridge, Mass: Harvard University Press.

Guala, Francesco. 2005. *The Methodology of Experimental Economics*. Cambridge University Press: Cambridge.

Hempel, Carl G. 1945. "Studies in the Logic of Confirmation (I.)." *Mind* 54 (213): 1–26.

Hume, David. (1748) 2007. *An Enquiry Concerning Human Understanding*. London: A. Millar. Reprint, Oxford; New York: Oxford University Press. Citations refer to OUP edition.

Illari, Phyllis McKay, and Jon Williamson. 2012. "What Is a Mechanism? Thinking about Mechanisms across the Sciences." *European Journal for Philosophy of Science* 2 (1): 119–35.

Illari, Phyllis, and Federica Russo. 2014. *Causality: Philosophical Theory Meets Scientific Practice*. 1 edition. Oxford University Press.

James, Alan. 2001. "Catastrophic Failure of a Raise Boring Machine during Underground Reaming Operations." In *Failure Analysis Case Studies II*, edited by D.R.H. Jones, 159-168. Amsterdam: Pergamon, 2001. Originally published in D.R.H. Jones, *Engineering Failure Analysis* 4(1): 71-80 (Amsterdam: Pergamon, 1997).

Jimenez-Buedo, Maria and Luis Miller. 2009. "Experiments in the social sciences: the relationship between external and internal validity." In *SPSP 2009: Society for Philosophy of Science in Practice (Minnesota, June 18-20, 2009)(PhilSci Archive).*

Jones, D.R.H., ed. 2001. *Failure Analysis Case Studies II*. 1 edition. Amsterdam; New York: Pergamon.

Kroes, Peter. 2009a. "Introduction to Part III." In Vol. 9, *Handbook of the Philosophy of Science: Philosophy of Technology and Engineering Sciences*, ed. Dov Gabbay, Paul Thagard, John Woods, and Anthonie WM Meijers, 405 – 408. Amsterdam: Elsevier.

———. 2009b. "Foundational Issues of Engineering Design." In Vol. 9, *Handbook of the Philosophy of Science: Philosophy of Technology and Engineering Sciences*, ed. Dov Gabbay, Paul Thagard, John Woods, and Anthonie WM Meijers, 513 – 541. Amsterdam: Elsevier. Kroes, Peter, Maarten Franssen and Louis Bucciarelli. 2009. "Rationality in Design." In Vol. 9, *Handbook of the Philosophy of Science: Philosophy of Technology and Engineering Sciences*, ed. Dov Gabbay, Paul Thagard, John Woods, and Anthonie WM Meijers, 565 – 600. Amsterdam: Elsevier.

Leonelli, Sabina. 2010. "Packaging Data for Re-Use: Databases in Model Organism Biology." In *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*, ed. Peter Howlett and Mary S. Morgan, 325-348. New York: Cambridge University Press.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science* 67 (1): 1–25.

Maher, Patrick. 1999. "Inductive Logic and the Ravens Paradox." *Philosophy of Science* 66 (1): 50–70.

Mill, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. John W. Parker.

Nightingale, Paul. 2009. “Tacit Knowledge and Engineering Design.” In Vol. 9, *Handbook of the Philosophy of Science: Philosophy of Technology and Engineering Sciences*, ed. Dov Gabbay, Paul Thagard, John Woods, and Anthonie WM Meijers, 350 – 374. Amsterdam: Elsevier.

Norton, John D. 2003. “A Material Theory of Induction.” *Philosophy of Science* 70 (4): 647–70.

Pearl, Judea and Elias Bareinboim. 2014. “External validity: From do-calculus to transportability across populations.” *Statistical Science* 29 (4): 579 – 595.

Peirce, Charles Sanders. 1883. “A Theory of Probable Inference.” In *Studies in Logic by Members of the Johns Hopkins University*, 126–81. New York, NY, US: Little, Brown and Co.

Petroski, Henry. 2001. “Success and Failure in Engineering.” *Practical Failure Analysis* 1 (5): 8–15.

Russell, Bertrand. (1912) 1997. *The Problems of Philosophy*. London: Home University Library. Reprint, New York: Oxford University Press. Translated by John Perry. Citations refer to OUP edition.

Russo, Federica, and Jon Williamson. 2007. “Interpreting Causality in the Health Sciences.” *International Studies in the Philosophy of Science* 21 (2): 157–70.

Slaughter, Robert H. Jr., Peter T. Cariveau, and Vincent W. Shotton. 2006. Back reaming tool. US7137460 B2, filed March 17, 2004, and issued November 21, 2006.

Steel, Daniel. 2007. *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford ; New York: Oxford University Press.

Talesnick, Mark, and Rafael Baker. 1998. “Failure of a Flexible Pipe with a Concrete Liner.” In *Failure Analysis Case Studies II*, edited by D.R.H. Jones, 31-43. Amsterdam: Pergamon, 2001. Originally published in D.R.H. Jones, *Engineering Failure Analysis* 5(3): 247-259 (Amsterdam: Pergamon, 1998).

Vickers, John. 2014. “The Problem of Induction.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2014.