Author's Preprint
See DOI 10.1007/s11229-014-0571-6 for version published by Synthese

Critical rationalism and engineering: methodology

Mark Staples

Accepted 1 October 2014

Abstract Engineering deals with different problem situations than science, and theories in engineering are different to theories in science. So, the growth of knowledge in engineering is also different to that in science. Nonetheless, methodological issues in engineering epistemology can be explored by adapting frameworks already established in the philosophy of science. In this paper I use critical rationalism and Popper's three worlds framework to investigate error elimination and the growth of knowledge in engineering. I discuss engineering failure arising from the falsification of engineering theories, and present taxonomies of the sources of falsification and responses to falsification in engineering. From this I discuss contexts of research and design in engineering, ad hoc rescue of engineering theories, and engineering assurance.

Keywords Engineering · Methodology · Falsification · Critical rationalism

1 Introduction

Engineering works within different problem situations than science. The specific activities of scientists and engineers have similarities: both develop and probe descriptive theories about the physical world, and both design and build things. Nonetheless, the ultimate problem situations are different[1]. Broadly,

M. Staples
NICTA, Level 4, 223 Anzac Parade, Kensington, NSW 2052, Australia
e-mail: mark.staples@nicta.com.au

M. Staples
School of Computer Science and Engineering, University of New South Wales,
Sydney, NSW 2052, Australia

[1] Brooks (1996, p. 62) says of this: "A high-energy physicist may easily spend most of his time building his apparatus; a spacecraft engineer may easily spend most of his time studying the behavior of materials in vacuum. Nevertheless, the scientist *builds in order to study*; the engineer *studies in order to build*."

science tries to understand the world, whereas engineering tries to change it. So, engineering has its own kind of knowledge which is similar but different to knowledge in science. Empirical theories in engineering concern not just the behaviour of artefacts, but also specifications of requirements for use, designs for artefacts, and claims that artefacts meet requirements for use. (Staples 2014)

Engineering also has its own ways of growing knowledge, which are again similar but different to those in science. Engineering epistemology can be explored by adapting frameworks already established in the philosophy of science. In this paper I use the perspective of critical rationalism, and build on an ontological basis developed in a previous paper (Staples 2014) to explore methodological issues in the growth of engineering knowledge. The previous paper discussed problem situations and the nature of tentative theories in engineering; this paper examines error elimination. Critical rationalism is not proposed as a perfect model of epistemology for engineering, but is instead used because it is a well-known and simple model that is adequate to represent and explore many aspects of engineering epistemology.

In section 2, I first recap the models of engineering knowledge presented in the previous paper, and then argue that (contrary to earlier claims in the literature) engineering theories are not inherently false. Nonetheless, they are falsifiable. Engineering theories make claims that artefacts meet requirements for use, and so falsification in engineering occurs when artefacts fail to satisfy requirements in usage situations, despite the predictions of engineering theories. This is engineering failure. In section 3 I use my ontological framework as a taxonomy of falsification of engineering theories: artefacts might not correspond to designs, engineering theories linking designs to requirements specifications might be invalid, or requirements specifications might not correspond to requirements for usage situations. Section 4 describes the various possible responses to falsification in engineering: to change requirements (or requirements specifications), to change artefacts (or designs), or to change engineering theories. I discuss the issue of whether in engineering the *ad hoc* rescue of theories is legitimate. I then draw together the discussion of falsification in engineering using critical rationalism's schema for the growth of knowledge. Because the scope of engineering theories is highly variable, responses to falsification can be made at different levels of generality, ranging on a continuum from a research context of generic problem-solving through to a design context for the design and analysis of specific artefacts. Finally I discuss the importance of assurance in engineering, and whether critical rationalism can support explicit justifications for assurances.

2 Engineering Knowledge

There are three recurring elements in definitions of engineering: physical artefacts (usually designed, but not necessarily so); requirements for these artefacts; and engineering theories for predicting whether artefacts will satisfy

those requirements. 'Artefacts' here include not just purpose-built tools, but any configurations of matter and energy used for some purpose. This can include structures, devices, system components, systems, and physical processes. Artefacts may have been originally designed with their current purpose for use in mind, but may instead have been re-purposed for a different kind of use, or may have emerged naturally with no original purpose (e.g. mine sites, or natural water sources used for cooling). This conception of artefacts is called 'useful-material' by Houkes and Vermaas (2009).

Requirements encompass not just the main function for the use of an artefact, but also acceptable side-effects, operational constraints, and environmental conditions. Rapp (1981) notes that while a technological action may achieve its main goals, it may nonetheless be abandoned if its unavoidable side-effects or secondary consequences are unacceptable. The importance of specific side-conditions and their acceptable range are often initially not known until they have been violated in use.

2.1 A Model of Engineering Knowledge

I previously presented a model of objective engineering knowledge (Staples 2014). Such knowledge is communicated through written text, diagrams, and mathematics, via university curricula, textbooks, patents, technical reports, journal papers, conference presentations, and technical standards. Objective engineering knowledge includes the designs of artefacts and specifications of requirements, but also includes engineering theories used to analyse artefacts and evaluate whether they meet their requirements. I define 'engineering theories' as explicit falsifiable claims used to predict and analyse the performance of artefacts with respect to requirements. Engineering may apply scientific theories, but engineering is not 'applied science' (Layton 1971; van de Poel 2010; Staples 2014). Engineering theories include not just general and fundamental theories taken or adapted from science, but also limited practice-based empirical claims represented by performance data tables, safety factors, and schematic diagrams.

My ontological model draws on Popper's (1972; 1977; 1978) *three worlds* framework. World 1 is the world of physical entities and phenomena. World 2 is the world of mental or subjective states and events. World 3 is the world of objective knowledge: abstract objects of thought that can be explicitly recorded or spoken. My model abuts two instances of Popper's framework. One instance represents the relationship between requirements on usage situations (World 1), and their formalisation as requirements specifications (World 3). The other instance represents the relationship between artefacts' presence and performance in the world (World 1), and their formalisation as designs (World 3). The two instances of Popper's framework are connected by reasoning within World 3 about why the predicted behaviours of designs will meet specified requirements, and therefore (conjecturally, according to an engineering theory) why actual artefacts (when those artefacts correspond to

the designs) will perform in use as required (when those requirements for use correspond to the specified requirements).

Within this framework, an engineering theory may play one or more of five roles, each of which supports claims about relationships between elements in the three worlds:

1. characterisations of changes required to be brought about in usage situations, represented (i.e., formalised, documented) as requirements specifications;
2. claims about how high-level requirements specifications can be decomposed to lower-level requirements specifications;
3. claims about how requirements specifications can be satisfied by a design;
4. claims about how high-level designs can be decomposed to lower-level designs;
5. characterisations of artefacts and their behaviour, represented (i.e., formalised, documented) as designs and descriptions of behaviour.

The first and last roles capture relationships between World 1 and World 3. The other roles capture relationships within World 3. Role 3 includes what Polanyi (1958, p. 176) calls 'operational principles'.

Engineering theories can be seen abstractly in logical terms, analogous to the deductive-nomological view of scientific theories. A general logical form for engineering theories is as follows:

$$[E(x,a); D(a)] \vdash R(x,a) \tag{1}$$

That is, an engineering theory claims that for any state of the world x, including an artefact a, where environmental conditions E apply in the world and to the artefact, and where the artefact fits a design D, then requirements R will be satisfied. When applied to reason about a specific artefact, its requirements must be within its predicted performance, and limitations on side-effects, operating constraints and environmental limitations must be allowed by the environmental conditions E.

The formula above can be decomposed using *modus ponens*:

$$[E(x,a); D(a)] \vdash B(x,a) \tag{2}$$

$$[E(x,a); D(a)] \vdash B(x,a) \rightarrow R(x,a) \tag{3}$$

That is, engineers may use one set of theories (formula 2) to predict artefacts' behavior B, then separately reason (formula 3) about how behavior of that kind satisfies requirements R. This problem decomposition allows the use of more generic theories to predict performance. The claims of a general theory are unlikely to be identical to particular requirements specifications R, but may entail them.

In terms of the roles played by engineering theories, formula 2 plays role 5 (artefact's behavior), and formula 3 plays role 3 (requirements satisfied by a design). Formulae 1 and 3 also play role 1 (requirements as specified), by

specifying requirements R as the goal. Theories of roles 2 and 4 are not illustrated here, but allow for the (recursive) decompositions of formulae 2 and 3 respectively. A logical perspective on design decomposition in engineering is discussed further by Hoare (1996). Rushby (2013) also provides a logical treatment of abstract requirement specifications, in the context of safety case arguments. Taken together, these serve as examples of the logical formulation of engineering theories.

2.2 Engineering Theories are not Inherently False

Popper (1959) says that the first characteristic of a theory is that it should be logically consistent, i.e., not be in contradiction with itself. Inconsistent theories can derive anything, including false conclusions. Inconsistent empirical theories can thus predict that impossible phenomena will occur and that certain phenomena will not occur, and so are worthless to science. An internally consistent empirical theory can only be falsified by empirical test. Observations when formalised become new givens (axiomatic) within the logic, and may then lead to a contradiction with a prediction derived from the theory, falsifying this new theory (i.e., the old theory plus observation), and leading to its rejection.[2] Just as it is irrational for scientists to use a theory known to be false to reason about the world, so too it is irrational for engineers to use a theory known to be false to reason about artefacts in the world. Both science and engineering are based on, and require, rational reasoning. Engineers that use theories known to be false will end up with undesired failure of their artefacts in use, when the usage situation is a false-valued instantiation of the theory. For safety-critical systems this will likely lead to deaths when those systems are used.

However, it is sometimes thought that engineers use 'false' theories. For example, Newton's theory of gravity has been falsified by observations of Mercury's orbit, but it is still used widely by engineers for many purposes. Popper (1963, p. 74) says "most formulae used in engineering ... are known to be false". I argue that a better view is as discussed in the previous paper (Staples 2014)—engineering theories are not necessarily false, but instead typically have a limited scope of applicability, and make approximate claims. To be valid from a logical perspective, an empirical theory (in science or engineering) must be true under all observable interpretations. Empirical theories can carry side-conditions that represent limitations in scope or precision without being inherently false.[3] Popper (1963, p. 74) does recognise that 'false' theories "... may be excellent approximations". An approximate theory may not be

[2] The appropriate response in this situation is not always initially clear, but ultimately one must either reject the general theory, or reject the formalisation of the new observation.

[3] Of course, a specific theory may turn out to be false, in which case it is rejected and a new tentative theory may be proposed to replace it. See section 4.1 for a discussion of *ad hoc* rescue of theories.

enough for scientific purposes, but can be good enough for many engineering purposes.

Wimsatt (2007) describes ways in which a scientific model can be 'false': having (limited) applicability, being approximate, or being merely phenomenological. A phenomenological theory is one which predicts phenomena, but whose elements are not intended to correspond ontologically to real entities in the world. These reasons may make a theory unacceptable in science, but as discussed previously (Staples 2014) they are not sufficient reasons to reject an engineering theory, and do not make a theory logically or empirically false. Wimstatt also identifies some ways in which a model can be empirically false: being incomplete (leaving out influential variables), revealing spurious correlations, being 'wrong-headed', or failing to correspond to the data. These are all problems that can lead to contradiction of an empirical theory with observation and are dealt with by rejecting the theory and perhaps revising it.

Cartwright (1983) says that fundamental scientific laws, in seeking to be explanatory, fail to correspond to facts. This is because such laws separate the component causes of physical situations. The laws either rely on *ceteris paribus* conditions which if explicated and taken literally, limit the scope of the laws to unrealistic situations that fit their ideals, or else rely on an essentially ad-hoc combination with other fundamental laws to explain any specific situation. So, Cartwright argues that for empirical theories, there is a trade-off between explanatory power and descriptive adequacy. She notes in contrast that engineering focuses on descriptive adequacy at the expense of explanatory power. This is amplified by Pirtle (2010) who says engineering theories may thus be better at telling the truth than fundamental scientific theories!

2.3 Falsification of Engineering Theories and Engineering Failure

I have defined engineering theories as supporting claims that artefacts perform according to their requirements. An engineering theory is falsified when its claims do not hold, i.e. when artefacts do not meet their requirements for use as predicted. Falsification in engineering can be identified by what is more commonly known as engineering 'failure'.[4] Engineering failures can be infamous, such as with the Tacoma bridge collapse, Therac-25 radiation overdoses, or mining disasters. An impact of less dramatic engineering failures is seen by the public whenever a product recall is made on a household appliance. Engineering failures can arise not just from physical failures of artefacts, but also from failures of theories to identify flaws in a bad design that have led to the artefacts being used, or from failures of the formalisation of a requirement (including environmental assumptions) to correspond to real

[4] Not every artefact failure will falsify an engineering theory. Often, acceptable requirements for use are qualified by probabilistic reliability conditions. This allows the use of imprecise engineering theories that accommodate occasional failure, arising for example from inevitable variations in quality of materials used in the construction of artefacts.

requirements. A design may be predicted to fail according to engineering theories, or the artefact may fail empirically in test or operation. The importance in engineering of identifying and avoiding failure is widely acknowledged, and the analysis of failures is critical to the growth of knowledge in engineering. Petroski (1996, p. 90) says that in engineering, "obviating failure is always the underlying principle". The requirements for an artefact in part comprise failure criteria, which define allowable limits on the performance of the artefact. Johnson (2009) notes that during the evolution of the design of automotive antilock braking systems that [p. 16] "...the knowledge produced by failed design was developmentally critical, as it led to new ways of thinking about the design of antiskid devices." This growth of knowledge in braking systems was not just in the design of those systems, but also in the understanding and characterisation of the requirements for those systems.

3 A Taxonomy of Falsification in Engineering

Petroski (2012, p. 360) says "the best way of achieving lasting success is by more fully understanding failure." A chain of beliefs, evidence, reasons, and actions lead to artefacts being used to address requirements. Engineering failure can arise from one or many failures in this chain. Engineering analyses work by trying to identify and remove such failures, so preventing the futile or dangerous use of bad artefacts.[5] Correcting errors in engineering knowledge relies on understanding the nature of those errors.

Overall, in terms of the three worlds ontological model, the failure of an artefact to satisfy requirements is a failure of the correspondence of a relationship between World 1 entities:

Usage Situation vs. Artefact as Built
: The behavior of the actual artefact in use fails to satisfy its requirements for use.

This can be decomposed by analysing the possible causes of falsification in terms of failures in the correspondences between various Worlds in the three worlds ontology of engineering knowledge. The first group of correspondences I consider concern engineering theories.

[5] Failures can be false negatives (where a bad artefact is mistakenly not identified as such) or false positives (where a good artefact is mistakenly thought to be bad). A false negative during design will not necessarily lead to a failure in use, because the artefact may satisfy its requirements in all of the specific situations in which it actually used, even though it would not function correctly in all of the specific situations it was required to be able to be used. False positives do not typically lead directly to engineering failures because the artefacts are in reality good, and as the artefacts are incorrectly deemed unsuitable, they are typically not used anyway. As well as false negatives and false positives, engineering theories may simply fail to show whether an artefact will meet its requirements. That is, theories may be indeterminate in some situations. However, like false-positive situations, indeterminate analyses do not lead to artefacts being deemed suitable for use, and so typically do not directly lead to engineering failures. In this paper I mostly focus on false-negative situations.

Requirements Specification vs. Usage Situation

A requirements specification may fail to capture the changes that are required to be brought about in usage situations. In logical terms, this is represented by falsification of formulae 1, i.e. by theories playing role 1. The specification of needed uses include not just the main functions of the artefact, but also allowable side-effects, and acceptable environmental and operational constraints.

In Popper's three worlds model, the connection between Worlds 1 and 3 is mediated by World 2. So, requirements specifications (World 3) may fail to agree with usage situations (World 1) because either of them have been misunderstood (in World 2). Examples of such misunderstandings are discussed below. Usage situations can also change. Often, an artefact which meets previously-unmet requirements places its users into a new physical situation in which there are new or different requirements. At an extreme, this can be a kind of 'wicked' problem (Rittel 1972), where outcomes are discontinuous, emergent, and impossible to predict. Emergent requirements may not be predictable in advance of the existence and use of the artefact.

Design Specification vs. Artefact as Built

An artefact might not behave as was predicted by the analysis of its design. In logical terms, this is represented by falsification of formula 2, i.e. by theories playing role 5 (or 4). This situation is very similar to normal empirical falsification in science. As in science, acknowledging the Duhem-Quine problem, there are many ways it can arise. The theory used to analyse the design may be wrong, or the observation of its performance may be invalid, but perhaps more commonly the artefact is merely built incorrectly. If a design is misunderstood by a builder, then it is unlikely to correspond with the built artefact. However, even if the design is well-understood by the builder, there may be a physical mistake made in the construction of the artefact, either as a 'slip' in its assembly, or through the materials used not conforming to their required quality. Examples of both are given below.

Design validity is a question of whether the design corresponds to the real system, but it is not always the case that the system is built according to that design. Sometimes engineers are given a fixed artefact to work with, either because they are recovering a lost design for an existing artefact, repurposing an existing created artefact, or working with a naturally occurring artefact (such as a mine site). In such situations, this correspondence is a failure of the 'design' to validly represent the artefact. Here engineering theories are used descriptively, more like scientific theories. Though, unlike scientific theories, it is sufficient for the engineering theories to be phenomenological. (Staples 2014)

Requirements Specification vs. Design Specification

The predicted behavior of the design for an artefact may fail to support the specified requirements for the artefact. In logical terms, this is represented by falsification of formulae 3, of theories playing role 3. This correspondence failure is of relations between World 3 objects: the formalisations of

requirements and designs, and the engineering theories that are used to reason about those formalisations.

The most commonplace occurrence of this failure is fundamental to the engineering design process. A tentatively-proposed design can be rejected if design-time theoretical analysis shows that it will not lead to an artefact that satisfies its requirements. This is not a false negative but instead a true positive—assuming the engineering theories are valid, the design itself will be flawed. The process continues by considering a new alternative design, perhaps a revised variant of the flawed design.

However, there are other potential causes. In practice it is possible to make calculational mistakes in the use of a theory, or to use an engineering theory whose side-conditions are not satisfied by constraints defined in the requirements specification or the design. In the logical view of engineering theories, these are all different kinds of unsound deduction, which could lead to either false-positive or false-negative conclusions.

Designs need not be exhaustive descriptions of every characteristic of an artefact. They only need capture aspects of the artefact that are important to satisfy the requirements. However, sometimes design abstractions can ignore aspects of the real world which lead to requirements being violated. Successful attacks on security-critical systems (i.e. failures of those systems) often work by breaking the validity of designs in this way. For example, 'TEMPEST' attacks on cryptographic communication works by observing stray messages on 'side channels' that reflect the original unencrypted message. (Anderson 2008, chpt. 17) These channels can include electromagnetic emissions unintentionally leaked by the input device or encryption device. Early designs for such systems did not constrain these emissions, and ignored the possibility of side channels being present. Side channels would not be important if the only requirement for the communications system was to deliver messages to the recipient, and so if present would not make the design invalid. However, if there is a security requirement that messages be kept confidential, then side channels can work to break the validity of the design with respect to this requirement.[6]

Artefact behaviours are not necessarily directly commensurable with requirements. Different engineering theories may be used for analysing artefact behavior and analysing requirements satisfaction. In practice, engineer-

[6] Some readers may wonder why an example from computing is relevant to a paper on engineering. The short answer is that computer systems engineering (including software engineering) is, from a methodological perspective, part of engineering. As argued in the previous paper (Staples 2014), computer programs as written embody a kind of objective knowledge, in World 3. Nonetheless, computer programs, when executing, are physical processes in World 1. They execute on physical hardware, and their execution leads to physical changes in usage situations. Indeed the example of TEMPEST attacks strongly supports the position that computer systems (combining hardware and software) are engineered systems, because the attacks exploit physical characteristics of the computer systems. Software as a formal entity in computer science is the wrong category of thing to be subject to physical side-channel attacks. Engineering theories about the security of software-based systems must include operational conditions and constraints on hardware in order to avoid falsification by such attacks.

ing theories are not usually explicated as logical formulae, and so partly-informal theories are used to bridge requirements and designs. Key assumptions or constraints are often left implicit. The use of such quasi-theories can lead to invalid conclusions and ultimately to engineering failures.

As well as failures of the correspondence between World 1 and World 3, there are potential failures related to understanding or realising artefacts or requirements, i.e. involving World 2. These are briefly discussed below.

Usage Situation vs. Understood Requirements
The physical changes required in a usage situation may be misunderstood. This includes the main functional requirements, limits on acceptable side-effects, and acceptable operational or environmental constraints. These misunderstandings are a common source of engineering failure. Often users do not know what they want. Instead of identifying an underlying need, users may mistakenly think that a superficial aspect of the usage situation drives their requirements. For example, a user may when driving a car, recognise that it is difficult to manually maintain a specific speed by keeping their foot steady on the accelerator pedal, and so desire a way to maintain constant force on the pedal or throttle cable. Some early technologies for this problem worked by clamping the throttle cable. Meeting such a requirement would satisfy the user's need while driving on flat roads, but would not meet the user's broader need for constant speed when driving up or down hills, where the throttle must vary to maintain a set speed. Later cruise control technologies use closed-loop control with information about actual speed used to modulate the throttle. Still, such a system does not satisfy users' overarching need for safety, for example in situations where a car must slow down or stop to avoid collision. Recognising that a problem exists in a usage situation does not mean that users or designers will be able to understand (let alone articulate) the nature of that problem, nor their requirements in the usage situation. There are a variety of engineering techniques that attempt to address this issue, including task analysis, prototypes, walk-throughs, and user-experience studies. In the broad development of technologies, the growth of understanding about requirements is tied up with the Lakatosian process for engineering requirements, as discussed below.

Understood Requirements vs. Requirements Specification
Well-understood requirements may not be completely or correctly formalised as requirements specifications. Requirements specifications can be documented using natural languages, modeling notations, or mathematics. These are complex kinds of objective knowledge, and may require significant expertise and study to be fully compared with one's understanding of the real requirements. It is not just the immediate statement of the requirements specification which must be understood, but also the theoretical consequences of the specification under applicable engineering theories. Failure to completely appreciate these consequences can lead to Lakatosian

'monsters' in requirements specification, as discussed further below. The case of 'System Z' discussed there is an example of this.

Design Specification vs. Understanding of Performance

When a design is used prescriptively (to inform the creation of a new artefact), then the design is given, and the artefact's behavior must correspond to the design. However, although the design may be good, the builder may arrive at a incorrect understanding of it, potentially leading to the construction of a bad artefact. An example of this is the fatal collapse of the suspended walkways in the Hyatt Regency Kansas City, in 1981. (Marshall et al. 1982) The design showed single rods supporting two walkways, but during construction, pairs of rods were used instead, which placed twice the load on supporting nuts on the higher walkway. This was a misunderstanding or misjudgement about the equivalence of the two solutions.

When a design is used descriptively (for example, during design recovery, or for a naturally-occurring artefact), the artefact is given, and the design must correspond to the artefact's behaviour. However, although the understanding of the artefact's behavior may be valid, the created design may be misunderstood and not be completely or correctly documented in the design.

Understanding of Performance vs. Artefact as Built

When a design is used prescriptively, then even if the design is well understood, the artefact may be constructed incorrectly. This may be due to a mistake (slip) in construction, or to faults in the materials used. An example of this is the collapse of the Kings Bridge, Melbourne in 1962 (Barber et al. 1963), which failed because of joints which had initially cracked immediately after welding.

When a design is used descriptively, then there may be invalid or incomplete observations of the performance of the given artefact, leading to an inaccurate understanding of its performance.

3.1 'Lakatosian' Falsification of Requirements

Here I amplify an observation by MacKenzie (2001) to argue that the growth of knowledge about engineering requirements resembles the process of *Proofs and Refutations* for mathematics (Lakatos 1976). In mathematics, tentative proofs are sometimes 'refuted' not by logical contradiction (in World 3), but instead by a logically-sound consequence (in World 3) that is unexpected and unwanted (in World 2). Lakatos argues that in such situations mathematicians should not just implicitly reject the unwelcome result ('monster barring'), but should instead either change their acceptance of the result ('concept stretching'—this is a response in World 2), or change the statement of the conjecture either by explicitly incorporating a condition that disallows the unwanted consequence ('lemma incorporation'—this is a response in World 3), or by inventing a wholly new conjecture.

MacKenzie (2001) has proposed that there has been a 'Lakatosian' process in the evolution of the understanding of requirements for the security of computer systems. The formalisation of the established computer security policy model (Bell and LaPadula 1973) was challenged by McLean (1985), who proposed a counter-example ('System Z') which satisfied the logical conditions of the formalisation but allowed the undesirable practical consequence of letting any user access any file. There was heated debate within the field on whether this example did constitute a counter-example. MacKenzie notes a range of Lakatosian responses to the counter-example that were seen in the literature, including attempts at monster barring and 'monster adjustment' (a kind of concept stretching, where the counter-example's consequences are accepted as a potential benefit), and ultimately a reconceptualisation and reformulation of requirements for computer security. This example is Lakatosian in a straightforward way, because the System Z counter-example was derived formally, based on a formal description of the Bell-LaPadula security policy model (all in World 3). No counter-example artefact was physically built and tested. The World 2 judgement was made about a World 3 result, and about the potential World 1 consequences should such a system be built.

Vincenti (1990) discusses another example of the evolution of knowledge of requirements: flying qualities. These are the kinds of response made by an aircraft in response to actions by pilots on the aircraft's controls. Flying qualities are largely an inherent aerodynamic characteristic of aircraft, but are experienced and evaluated by human pilots. In the 1910s there was an awareness of a trade-off between stability and control in the design of aircraft, and that both impacted flying qualities. Engineers could distinguish between static stability (the tendency to return to original level flight after a disturbance) and dynamic stability (the tendency for oscillations around the original attitude to progressively dampen). A theoretical treatment of stability in 1911 dismissed the importance of short-period dynamic stability for flying qualities, focusing instead on static stability and long-period dynamic stability. By the 1920s, designers in practice were also ignoring dynamic stability. Vincenti says [p. 73] that by 1935, it had become "obvious" that designers could target flying qualities for different kinds of aircraft by varying static stability, though theoretical studies and textbooks continued to provide analytical treatments of long-period dynamic stability. The underlying requirement for use was that planes flew well, but the requirement specifications were stated in terms of conditions on stability, including long-period dynamic stability, but not short-period dynamic stability. However, during the 1930s, although new aircraft designs met these requirement specifications, they continued to display unexpected flying qualities—in practice it remained unclear how to avoid these problems and achieve good flying qualities in use.

The resolution of this problem depended on better understanding the requirements. Empirical research began to explicitly compare stability characteristics and quantitative measures of the experience of the pilot in the cockpit during flight, including measures of joystick forces. When comparing stability to more detailed measurements of actual experience of the pilot, the experi-

ments [p. 79] "... brought into doubt the twenty-five-year preoccupation with the long-period oscillation, at least so far as flying qualities were concerned." Despite the established theory, long-period dynamic stability was not related to pilot's perception of flying quality. The unquestioned assumption that short-period dynamic stability was unrelated to flying qualities then also began to be re-examined experimentally, with reference to the new understanding of flying qualities. The requirements for use were grounded in the experience of pilots. Requirements specifications for flying qualities had originally been given partly in terms of long-period dynamic stability, but this was rejected by the community when in 1943, requirements specifications shifted to instead use constraints on short-period dynamic stability. Ultimately what was rejected by the community was not the judgments about flying qualities by pilots, but the theory that flying quality could be specified by static stability and long-period dynamic stability.

Vincenti does not identify the growth of knowledge about how to specify flying qualities as Lakatosian, but we can identify some Lakatosian themes in his story. The established theory that flying qualities were determined by static stability and dynamic stability led to the creation of monsters, aircraft that unexpectedly had poor flying qualities. During the period between 1911 and 1943, theoreticians and designers were in some sense in denial (monster barring) about the reality of how stability related to flying qualities as experienced by the pilot. Vincenti notes the reason why this issue was not confronted earlier: because aircraft designers were instead focussed on other more important dimensions of aircraft requirements, especially improving flying performance (such as speed, range, and efficiency). When designs improved to the point of diminishing returns on flight performance, then flying qualities became the next most important dimension for technological progress. This is a 'reverse salient' in Hughes' (1976, p. 646) terminology. Eventually the community developed a new idea of what flying qualities were (concept stretching), and developed new theories for how to specify and achieve those requirements.

A difference between Vincenti's and MacKenzie's examples is that in computer security, the counter-example of System Z arose as a formal consequence (World 3) of the requirements specification (a 'presumptive anomaly' in Constant's (1984) terminology), whereas the poor flying qualities for aircraft were experienced empirically (World 1). Both were falsified in a 'Lakatosian' manner, by a World 2 judgement that the unexpected consequences were unacceptable. Both cases were resolved by better understanding the nature of the requirements for use, and better understanding how to specify them. In both cases this lead to the development of new engineering theories about how designs could satisfy those new requirements specifications. Engineering theories of requirements are, somewhat like scientific theories, tempered by the reality of physical phenomena, but are also, somewhat like mathematical theories, tempered by what is desirable. As in mathematics, the methodologically-sound approach in engineering is not to implicitly reject the failure. Instead engineers should document the acceptance of the behaviour as a new requirements spec-

ification, and ensure that artefacts perform according to the newly-explicated requirements and prohibit the unwanted behaviour.

4 Responses to Falsification in Engineering

When an empirical theory is used descriptively, if there is a failure of the predictions of the theory to correspond to phenomena of the real world, then the theory is wrong. Scientific theories are descriptive, and so the response to falsification is to reject or change the theory, assuming predictions are logically sound with respect to the theory, and that observations are valid with respect to their definition as theoretical constructs and underlying measurement theory. The Duhem-Quine problem means that we do not necessarily know which part of our theory or observation is wrong. Challenging the observation is one possible response, but another is to reject and revise the theory.

When an empirical theory is used prescriptively, if there is a failure of its predictions to correspond to real phenomena, then (if the predictive theory is valid) the real world is wrong. As acknowledged by Constant (1999), the Duhem-Quine problem continues to be an issue in engineering. However its effect is amplified in engineering, because for a prescriptive application of theory the real world can be wrong, and also because there are additional empirical theories in play: those related to requirements. So in engineering there is a wider range of responses to falsification than in science. Below I use the three worlds model of engineering knowledge as a taxonomy of this range of responses. To respond to falsification, engineers change entities or relationships in one or more parts of the three worlds model of engineering knowledge to attempt to restore the correspondences between the Worlds. This is integral to error elimination in the growth of engineering knowledge.

Change Requirements for Use or Requirements Specification If an artefact fails to meet its requirements, then sometimes it is possible to change those requirements. This may mean simply modifying the requirements specifications to make them better describe the changes that are required to be brought about in usage situations.

However, sometimes it is possible to change the needed uses, either by changing people's judgements about usage situations, or by changing the usage situations so as to correspond to people's judgements. Rapp (1981, p. 62) says that "reflection of the goals explicitly or tacitly pursued ... can clarify which existing facts one accepts as simply given and unalterable when, in fact, their (gradual) modification is possible." Changing requirements may entail changing the main function of the artefact, or may only involve changing the allowable conditions of use. In engineering practice, the acceptance of requirements change may be achieved by renegotiation with individual customers or users, or by product marketing with broader groups of consumers.

In the most extreme case, requirements may need to be changed because they are discovered to be infeasible to realise. Rapp (1981, p. 61) says "... the

tension between the desired and the actually attainable is reduced to a tolerable level by adapting goals to what is feasible." Often it is not initially known whether requirements are feasible or not. Given this, Rapp is concerned that "... it would be wrong to demand complete correspondence between desired goals and existent means, since this would imply that only that 'goal' could be envisaged for which all the means of its realization are already available." Technology often advances by tackling problems with no known solution. However, the initially indeterminate feasibility of requirements should not by itself rule out demanding correspondence between goals and means in engineering. Just as scientific theories may be tentatively proposed prior to severe test, so too engineering requirements may be proposed prior to knowing whether they are feasible to realise.

Change Design or Artefact If an artefact does not meet its requirements, then perhaps the most obvious response is to try to change the artefact so that it does. Rapp (1981, p. 60) says that engineers may "... retain the goals and employ alternative means to achieve them." It may be that the design is sound but the artefact is not a valid realisation of the design, in which case the artefact may be changed. Or, it may be that the design is not sound, and so both the design and artefact are changed.

Laymon (1989) describes engineering responses to the 'projectibility problem', where an empirical theory cannot be applied reliably to a design because of practical considerations that invalidate important idealising assumptions in the theory. Where an engineering theory has been validated by use in the successful design and realisation of prior artefacts, Laymon notes that the most common response to the projectibility problem is to create new designs that deviate as little as possible from the previous successful designs. A more sophisticated version of this response is for engineers to argue on the basis of some other underlying theory that the differences between the new and previous designs are not significant with respect to their negative impact on required performance. (Laymon gives as an example design arguments like this both for and against the Tacoma bridge, prior to its failure.) More sophisticated again is an engineering response which builds mechanisms to bring reality closer to the theoretical ideal in the designed artefact, by tighter control of more variables. (Laymon gives as an example design features that reduce friction or lower mass in the design of a suspension system.)

Reject or Change the Analytical Engineering Theory If an artefact fails to meet its requirements despite the predictions of an engineering theory, but the corresponding design and requirements specification are valid, then the theory is invalid and must be rejected. (Here I assume that the artefact's failure is not allowable under reliability conditions present in the requirements.) Future theoretical analysis must use an another theory. This may be an entirely new engineering theory, an alternative pre-existing engineering theory, or a revised version of the rejected theory. The rejection of engineering theories and the development of new ones is how the growth of engineering knowledge occurs,

and is discussed below in section 5. The revision of a rejected engineering theory will sometimes occur by weakening the rejected theory, in particular by approximation (reducing claimed precision or accuracy), or by qualification (adding extra conditions). Concerns about the ad-hoc rescue of engineering theories in this manner are discussed below in section 4.1.

During the design process, an engineering theory may be indeterminate about whether an artefact will meet its requirements. This may happen because the artefact's performance is close to the limit of a particular requirement, and the engineering theory being used is not precise enough to distinguish whether the artefact will meet the requirement. In such situations, engineers will choose an alternative theory which has greater precision, but is perhaps more difficult or costly to use. Situations where there is no applicable precise-enough theory to analyse a particular design are the bailiwick of engineering research, discussed below in section 5.1.

Multiple Simultaneous Responses Often the response to engineering failure is a combination of the responses above. New requirements often lead to new designs, so that both the requirements and design are changed. Johnson (2009) provides examples of how knowledge about automotive antilock braking systems grew not just as evolving solutions (designs), but also simultaneously as evolving questions (requirements specifications). Many engineering theories are specific to a class of designs, and so if an alternative tentative design is chosen, a new matching engineering theory may also be needed to analyse the design. Similarly, when addressing stronger requirements, a more precise theory may need to be used.

4.1 Ad Hoc Rescue of Engineering Theories

From a logical point of view, the validity of an empirical theory that has been falsified by an empirical observation in particular circumstances can be immediately be 'rescued' by adding a caveat that the theory is not intended to be applicable in *those* circumstances. This is called the *ad hoc* rescue of theories. Although it is logically admissible, there are methodological concerns.

Popper (1959, 1963) objected to the *ad hoc* rescue of theories because of concerns about the demarcation and goals of science. Astrology, for example, while clearly not scientific, might repeatedly use this strategy to avoid holding to any prophecy that has just been falsified. Popper objects to an instrumental view of theories for science partly because he believed that instrumental theories can legitimately be rescued by *ad hoc* qualification. Popper claimed engineers use false theories (for example, Newton's theories) instrumentally, but then claimed (1963, p. 152) in a seeming contradiction, that instrumental theories cannot be falsified! Popper's seeming contradiction arises because falsifiability is more a methodological quality than a logical one. Popper often discussed science in terms of a single best currently accepted theory within

a field. Lakatos (1970, p. 188) presented the methodological nature of falsifiability more clearly by arguing that it is better to think about scientific method operating not over isolated theories, but instead over a series of theories. This view does not change the basic logical position, but it becomes easier to present the methodological issues. Lakatos (1970) maintained Popper's objection to the *ad hoc* rescue of theories in science. For science, Lakatos says not all qualifications are *ad hoc*. 'Progressive' problems shifts are OK: these are at least 'theoretically progressive' (by predicting new facts), and are ideally also 'empirically progressive' (by already having empirical corroboration). So, the rescue of scientific theories isn't always inherently illegitimate. In engineering too, rescue which merely disqualifies past counter-examples without constraining future predictions is not 'theoretically progressive' and remains methodologically unsound, because it does not lead to any substantive improvement in knowledge.

One reason for avoiding *ad hoc* rescue in science is that it can overly-reduce the scope of the theory. Ideally, scientific theories should be as general as possible, but cheap rescue may impose overly-strong applicability conditions. This concern is not necessarily as important in engineering, because engineering theories do not necessarily aim for extreme generality. Nonetheless repeated explicit qualification of a theory progressively makes it less universal and less simple to use. The growing complexity and reduced scope of applicability of a theory rescued by narrow qualifications will likely eventually see it fall in preference to another, equally corroborated, less restricted, and less complicated alternative theory. Or, as argued in the previous paper Staples (2014), if the narrow theory is not dominated by the new theory across all criteria of interest, it may survive as one of the multiple overlapping theories at the forefront of engineering knowledge.

There is another difference between science and engineering concerning the rescue of theories. Engineering theories make claims not just about the performance of artefacts, but also about how that performance meets requirements. Engineering theories may be falsified not just because reality does not agree with predictions about artefacts, but also because the specified requirements do not agree with needed changes in usage situations. As discussed in section 3.1, there can be a Lakatosian (1976) process in the growth of knowledge of engineering requirements. Just as Lakatos notes that 'lemma incorporation' can sometimes rescue a mathematical conjecture from a counter-example, so too qualifications may be introduced to rescue engineering theories from the failure of the requirements specification to correspond with real or intended requirements. In engineering terms these are new operational constraints or environmental limitations that weaken the theory. However, if the resulting theory cannot provide a justification that the artefact will meet its requirements, then alternative responses will be needed.

5 Growth of Knowledge in Engineering

The growth of knowledge in engineering can be seen to use a method of conjecture and refutation similar to that in science. Both fit Popper's general schema for critical rationalism:

$$PS_1 \rightarrow TT \rightarrow EE \rightarrow PS_2$$

In an initial problem situation (PS_1), the creation of a tentative theory (TT) is tempered by the identification and elimination of error (EE), leading to a new problem situation (PS_2).

As noted in the introduction, engineering works within different problem situations (PS) than science. Tentative theories (TT) in engineering are also different to those in science. (Staples 2014) Unlike science, it is sufficient for engineering theories to be phenomenological. And, although not a characteristic difference, engineering theories are also often narrower in scope and less precise than scientific theories. Of course, in some cases, such as for microchips, nanotechnology, or space technology, extremely high precision is required in engineering. Scientists value precision in its own right, whereas engineers value precision only to the extent that it helps artefacts to meet requirements. Because of this, the forefront of engineering knowledge consists of many overlapping theories, each making different trade-offs between scope, precision, accuracy, and ease-of-use. As discussed in this paper, error elimination (EE) in engineering is also different than in science. There is a wider range of responses to falsification in engineering than in science, because engineering theories can be used prescriptively (and so the real world may be wrong rather than the theory), and because claims in engineering concern not just the performance of artefacts, but also how those artefacts satisfy requirements.

The growth of knowledge is the process by which the best-available theories at the forefront of the body of knowledge become better over time. This includes adding knowledge about new things, and improving knowledge about existing things. Popper (1963, pp. 314–315) unsystematically lists six ways in which a scientific theory can be better than another:

1. has more precise assertions (that also withstand tests);
2. explains more;
3. describes the facts in more detail;
4. passes more tests;
5. suggests more previously-unconsidered experimental tests, and passes them; and
6. unifies or connects previously unrelated problems.

An engineering theory may be better than another in these same ways. The primary criterion for retaining an engineering theory in the body of knowledge is the correspondence of its predictions with observations of the real world. These predictions are not just of artefacts' behaviour, but also whether and how well this behavior meets requirements for use. We can regard the other criteria as being of secondary importance for engineering theories: being more broadly

applicable (having weaker antecedents), being more precise (having a stronger consequent), or having a better explanation in terms of other more fundamental theories. In engineering there are also additional secondary criteria such as being easier, less expensive, or faster to use as predictive instruments.

Creating explicit World-3 objective knowledge about requirements is critical to improving our understanding of requirements and to developing artefacts to meet those requirements. Rapp (1981, p. 62) says "explicit formulation and discussion of goals ... can reveal intentions which are otherwise only intuitively or implicitly assumed." This concerns not just the main goals for the artefact's requirements, but also knowledge about what logically-uninterpreted conditions must be given an explicit empirical interpretation to maintain the validity of engineering theories. Beyond the growth of knowledge about the formalisation of requirements is the growth of knowledge about what requirements are or should be. Rapp (1981) distinguishes as a kind of technological process the development of ideally desirable technology, where normative criteria are used and developed to assess whether realised technologies lead to a better world.

New or improved engineering theories can also concern improved designs or new kinds of designs for artefacts which are able to better meet requirements, or are able to meet new requirements. The design of an artefact is important not just to that artefact. Designs are objective content, and are part of the body of engineering theory. Ferguson (1992, p. 13) speaks of 'low-level inventing' in the creation of new designs for new artefacts, but based on prior well-tested principles. He contrasts this with 'fundamental invention': the creation of radical new 'operational principles' for whole classes of artefacts. Vincenti (1990, p. 208) similarly talks of 'fundamental design concepts' as an important category of engineering knowledge. Vincenti, like Ferguson, draws on Polanyi's concept of 'operational principle' for an artefact. This conceptualises how an artefact's (Polanyi 1958, p. 328) "... characteristic parts—its organs—fulfil their special function in combining to an overall operation which achieves the purpose". Vincenti [p. 208] expands on this by describing how 'normal configurations' ("... the general shape and arrangement that are commonly agreed to best embody the operational principle") specialise operational principles to together define technologies.

New or improved designs are the core of technological progress. Rapp (1981) says that technological progress is achieved when either the same work can be done with less input, or when higher output is achieved with the same input. The measure of input and output is not always straightforward, and can include and combine measures of qualities such as precision, speed, or reliability. Vincenti (1990) expands on anomalies identified by Constant (1984) to discuss three drivers for growth of design knowledge: 'presumptive anomaly' (theory-based predictions that the artefact will begin to fail to perform in future, or will be out-performed by an alternative), functional failure, and reduction of uncertainty in design. Laudan (1984a, pp. 84–85) similarly presents a taxonomy of technological problems, for which technological progress provides new or improved solutions:

1. a problem in the environment, unsolved by current technologies;
2. functional failure of a technology (e.g. when placed under new demands);
3. 'extrapolation from past technological successes';
4. imbalances between related technologies (including Hughes' (1976) 'reverse salients on a technological front'); and
5. potential (not actual) failures, predicted limits of technological performance (including Constant's (1984, p. 31) 'presumptive anomalies').

The growth of knowledge of new or improved designs is initiated by creative responses to these problems situations—the invention of theories (incorporating the invented designs) which tentatively claim that artefacts satisfying these designs will satisfy the requirements. These new design-specific theories may initially be tested using other more general and more well-established theories, but are ultimately tested by judgments about how well artefacts satisfy requirements for use. Successful designs give us greater power to change the world to our requirements, and add to the body of engineering knowledge.

5.1 Contexts of Research and Design in Engineering

Vincenti (1990, p. 226) says that scientists "use knowledge primarily to generate more knowledge", whereas engineers "use knowledge primarily to design, produce, and operate artefacts" but also "use knowledge to generate more knowledge". Thus "engineering knowledge has two uses instead of one". The application of engineering knowledge in the design and creation of specific artefacts is itself a place for the creation of hypotheses, empirical test, and the growth of engineering knowledge, if only within a narrow scope. Solutions for new design problems are rarely completely known before-hand. Even if the main functionality for a design problem is standardised and there are successful operational principles and normal configurations for the design, there can still be variation in operational and environmental constraints, or variation in the valuation of secondary requirements affecting trade-off decisions among design alternatives. For fully-standardised artefacts, design problems typically shift to questions about how to address more stringent cost-related requirements.

In engineering, we might consider there to be two contexts for the growth of engineering knowledge: a 'research context' of generic problem-solving leading to knowledge reusable across a number of engineering problems, and a 'design context' of creation and analysis of designs of specific artefacts leading to knowledge about the performance of those artefacts. Boon and Knuuttila (2009) somewhat-similarly distinguish *engineering* from *engineering sciences*, the latter being 'scientific' research about the behavior of devices, but the former being concerned with the creation of designs for devices. Vincenti (1990, p. 48) also discusses design knowledge at the artefact level (what should be the design for this artefact), and at the idea level (how does one come up with designs). Work in the research context might include the creation and test of new operational principles, new formulations of artefact requirements, new design methodologies, or new engineering theories for the analysis of artefact

behavior. Work in the design context might include the creation of individual designs, or the test and analysis of designs or prototypical models to explore relative performance of the artefacts in various performance dimensions. The schema of Popper's (1959) critical rationalist view of science is implicitly framed around the distinction between a context of discovery (for the creation of tentative theories) and a context of justification (for the test of tentative theories). The two engineering contexts of research and design discussed above are orthogonal to these two contexts of discovery and justification. The engineering contexts of research and design might each be considered to have their own contexts of discovery and justification.

There is not necessarily a sharp distinction between engineering research and engineering design. They are distinguished by the scope of supported functionality. Designs have varying levels of generality: some (such as for a mine site) may be highly specific to an individual physical situation, whereas highly parametric designs may apply in a regular way to a range of designs and thus to many more problem situations. Still more general design principles may apply to a broader swathe of designs and problem situations. Because there is no clear boundary along this continuum of design generality, I claim there is no clear boundary between the contexts of research and design in engineering.

The bulk of engineering practice concerns the design and analysis of specific artefacts. Nevertheless, there is knowledge creation within this context, concerning at least the specific new designs and their relationships with specific artefacts and requirements. In an empirical study of engineering practice, Gainsburg et al. (2010) found that 67% of structural engineering knowledge used in one engineering firm was practice-based. Practice-based knowledge can be more generally-applicable, within other companies in the same industrial domain. Johnson (2009) observes in her history of the development of automotive antilock that journal and conference publications included many significant articles by practitioners based within companies, and that this lead to greater technological progress within those companies.

In addition to the contexts of research and design, we might consider there to be other contexts for the growth of knowledge in engineering. For example, we might discuss the growth of knowledge arising during the *construction* (sometimes called 'production') of an artefact according to a design, or indeed in the *use* of the artefact (Kroes 2002). Vincenti (1990) devotes a chapter to describing engineering knowledge arising in the context of production. Knowledge certainly grows within the operational use of an artefact, as an empirical test of the engineering theories that led to the artefacts being used. Pols (2010) discusses the transfer of responsibility for artefacts from design through to use. In engineering, conditions arise through the use of specific theories used to analyse parts of a design, and are then carried by the design and artefact. If the conditions are not satisfied by an artefact being situated appropriately as part of a larger artefact or system, then ultimately the conditions will be transferred to use as specific operating conditions, or instructions for safe use. The user must be willing to accept these conditions of use as part of the acceptable requirements for the artefact, to ensure safe and efficient performance of the

main functional requirements of the artefact. 'Use plans', manuals, training, and other documentation are important to communicate these conditions to users, and to ensure that the conditions arising from designs and engineering theories used to analyse those designs are satisfied, so that the artefacts can function as designed.

Another important role for engineering knowledge in the context of use of artefacts is in accident investigation. Given an engineering failure to meet requirements (in particular, critical safety requirements), *modus tollens* allows us to deduce that one or more of the requirements, artefact or engineering theory is at fault, as described by the taxonomy of sources of falsification presented in section 3. Petroski (2012, p. 114) notes the difficulties in failure investigation: that "failure analyses are hypotheses heaped on hypotheses" but that (p. 237) "when a structure collapses, it is obviously no longer available in its pre-collapsed form for rigorously testing any hypotheses". Determining which is at fault relies first on the collection of diagnostic data about an accident, then on assembling and analysing that data, using engineering theories, in an attempt to diagnose likely causes of the accident. Of course, the particular engineering theories that led to the artefact being deemed suitable for use might potentially have been at fault, and so alternative engineering theories may be needed for the investigation.

5.2 Assurance in Engineering

Davis (2010) observes that architects design and builders build, but neither are taken to be engineers. Designing and building are important activities within engineering, but I claim that perhaps engineering's most important and distinctive role is to provide assurances that artefacts meet requirements for use. An assurance is a warrant provided on some basis by one person to another. People rely on engineering assurances to accept the use of artefacts to try to satisfy their requirements, and in particular to remain safe from the danger of death, injury, or property damage during the use of artefacts. The heavy responsibility of this role and the potential for individual and social impact from engineering failures is the reason why the engineering profession is legally protected and regulated in many countries. Engineering assurances are typically backed by explicit justifications that rely on the use of empirically-tested engineering theories. Often there is a legal requirement for an explicit and documented rationale, which can then be reviewed during an accident investigation or litigation.

Philosophical discussion has sometimes taken 'justification' to be an infallible argument for the universal timeless truth of some statement. Critical rationalism rejects the possibility of this kind of absolute justification for empirical claims. However, engineers are often required to provide justified assurance about the performance of artefacts. How then can critical rationalism deal with engineering assurance?

Moreover, different artefacts have different kinds of criticality, and more critical artefacts typically have higher assurance requirements. For example, artefacts can be safety-critical if their failure can lead to injury or death, can be security-critical if their failure can lead to the loss of sensitive information, or can be mission-critical or business-critical if their failure can lead to the failure of a project or company. All engineered artefacts come with some level of assurance, but safety-critical and security-critical artefacts typically require more stringent assurances. That is, they must be backed by stronger justifications. How are some claims more justified than others?

For most practicing engineers, engineering theories are regarded as suitable for use when they are accepted within the engineering profession. The acceptance of an engineering theory may be codified by national or international standards or certification bodies. Davis (2010) argues that the definition and acceptance of a core curriculum of engineering knowledge is key to the definition of the occupation of engineering as a profession. However, this only begs the question: how does the engineering profession gain enough confidence in new engineering knowledge for it to enter the curriculum, or for it to be recognised by a formal standard or certification body? And how can engineers develop sufficient confidence about entirely new technologies or analysis techniques? These questions above are key for a philosophy of engineering. Vermaas (2010) discusses criteria for justifiably ascribing a function to an artefact: that the artefact is justifiably believed to have a capacity to function in a way that leads to a user's goals when used appropriately, and that this belief is communicated appropriately to or by the user. However, these criteria circularly rely on a conception of 'justification' for belief in the functional capacity of the artefact.

The approach to take for engineering assurance should be similar to our approach for confidence in scientific theories. Tentative acceptance of engineering theories should depend primarily on them surviving severe empirical tests. Engineering justification must be provided using the best unfalsified theory which is available that supports the claim made that an artefact meets its requirements. However, there is an interesting twist to this issue in engineering. As discussed in the previous paper (Staples 2014), the forefront of engineering knowledge contains a multitude of overlapping theories, each of which may be 'best' in terms of a number of characteristics, including precision, accuracy, scope of applicability, and confidence. So this issue is more complex in engineering than in science.

Can we measure our confidence in an engineering assurance or engineering theory? Within engineering, a *probabilistic risk assessment* (PRA) is often performed in an attempt to quantify the confidence that engineers have in arguments for the safety of highly-critical systems. Engineers do not seek a probabilist quantification of their claims for its own sake. Regulatory or contractual requirements for such systems sometimes include requirements on the levels of confidence in a PRA; systems which are more critical require greater levels of confidence. Ongoing work in this area (Helton and Oberkampf 2004) attempts to provide explicit treatments of aleatory uncertainty (arising from

chance in the world) separate from epistemic uncertainty (arising from lack of knowledge about the world). For systems using redundant mechanisms to achieve very high reliability, epistemic uncertainty about artefacts or requirements can easily outweigh the aggregate aleatory uncertainty about the performance of components. However, are these PRA approaches philosophically credible? There is no deductively-valid way to argue on the basis of prior empirical evidence for the truth of empirical claims about the future of the world. But is there a rational probabilistic or stochastic basis for quantifying the strength of such claims? This is an open question.

Popper denied it was possible to provide infallible justification for theories, and was critical of any means for positive justification. Nonetheless he argued that tentative theories could be 'corroborated' by passing empirical tests. (Corroboration is not the same as absolute confirmation, because empirical theories are always held to be tentative.) Popper (1972) later attempted to sharpen this notion by defining 'verisimilitude' as a metric for the 'truthlikeness' of theories, i.e., how much one theory is closer to the truth than another. Such a metric might have helped to explain PRAs. However, Popper's attempt to define verisimilitude was unsuccessful (Oddie 1981), and it remains an open challenge. Popper's account of science in terms of deductive falsification never led to a satisfactory treatment of how, or the extent to which, theories were corroborated by surviving severe tests. So, critical rationalism does not provide an entirely satisfactory basis to investigate engineering assurance. Mayo (1996) has provided an alternative error-statistical philosophy of science, which may be more promising in this regard. Mayo does not intend her approach to be seen as a 'fix' for critical rationalism, but it is largely consistent with the overall schema of error elimination and improvement in critical rationalism, and may apply to the view of engineering methodology presented here. Although Mayo also rejects probabilism, she does allow (p. 446) that sometimes the correctness of a hypothesis can be interpreted as its reliability. This may help to explain aleatory uncertainty in PRAs, but it is unclear whether it can give a treatment of epistemic uncertainty.

Regardless of the basis on which an assurance was initially given, the rationality of a decision to provide assurance for an artefact must be judged with respect to the state of knowledge available at the time the decision was made, and with respect to the appropriate use of that knowledge to inform the decision. Of course artefacts with engineering assurances may later fail, because the artefacts do not correspond to their design, or because the best available theory turns out to be false. It is instructive to consider cases where engineering assurances have not been satisfied. A common kind of situation is consumer 'product recalls', where a product that was initially judged to be safe to use, may be later deemed to be unsafe, because evidence has emerged that the original engineering theory used as the basis for the justified assurance is shown to have no empirical validity. In some cases, this may happen even if the artefact has performed as required for many (or all!) of its users, and even if the artefact would in fact forever perform as required. Nonetheless, for

highly-critical artefacts, the loss of valid justification may be enough to force a recall, despite the absence of failures.

A final note: assurance is achieved not just through the use of engineering theories to analyse designs, but also concurrently through their use to shape designs. Clausen and Cantwell (2007) discuss how safety factors can be used both to provide conservative analyses of designs, and also to introduce conservative improvements to designs. However, higher levels of required assurance can sometimes be antagonistic with other performance characteristics. For example, the only way to demonstrate that an artefact is very strong or highly robust may be to make it undesirably heavier or more costly to manufacture because of the use of stronger or higher-quality materials. There is often a trade-off in engineering design between assurance and performance. This is one reason why engineers have continued importance not just in the evaluation of artefacts and designs, but also in their creation.

6 Conclusions

Modern engineering has produced technological advances that underpin much of humanity's power to change and improve the world. These developments derive from the growth of engineering knowledge. Engineering knowledge is important not just to make technologies work, but also to provide a basis for justified assurance for their performance and safety.

The first thesis of this paper is that because the nature of engineering is different to science, and theories in engineering are different to theories in science, so the growth of knowledge in engineering is different to the growth of knowledge in science. The second thesis is that methodological issues in the epistemology of engineering can be treated by adapting frameworks already established in the philosophy of science. I have used critical rationalism and Popper's three worlds framework, adapted as described in a previous paper (Staples 2014). The third thesis is that this can help us to understand the sources of error in engineering and responses to error in engineering, and that these are critical to error elimination and the growth of knowledge in engineering.

I have argued that, despite earlier claims in the literature, engineering theories are not inherently false. Instead, they use approximation and limited scope to retain valid phenomenological correspondence with the world. However, methodologically-sound engineering theories can be falsified. Falsification provides an opportunity to improve engineering theories, but effective improvement requires understanding the nature of the failure. I have used the three worlds model of engineering knowledge to provide a taxonomy of sources of falsification in engineering. Engineering theories can be used prescriptively or descriptively, and so if falsified, they may be wrong, or the real world may be wrong. Artefacts' performance may fail to correspond to the performance predicted by their designs, the predicted performance of designs may fail to satisfy requirements specifications, and requirements specifications may fail

to correspond with requirements on usage situations. Any of these sources of falsification can lead to engineering failure. The growth of knowledge about requirements is epistemologically interesting because the rejection of requirements can be similar to judgements about Lakatosian 'monsters' in mathematics. Sometimes people don't initially know what they want—the physical or formal consequences of requirements specifications may be unexpected and unwanted. The three worlds model also frames a taxonomy of responses to falsification. In engineering, responses to falsification are made to one or more of the sources of falsification in an attempt to restore the validity of correspondences.

Engineering knowledge grows in different ways than scientific knowledge. Much of the growth of engineering knowledge occurs through its use. Engineering research and engineering design practice can be viewed as two different contexts, but I have argued that they are distinguished largely by the scope of functionality. Engineering theories are used not just for design, but also for accident investigation, and to provide assurance that artefacts will meet their requirements, and be safe to use.

There are many epistemologically-interesting open questions about engineering. How should we understand changes to requirements arising from the use of artefacts? How does the epistemological nature of engineering relate to ethical issues in engineering? What underlies engineering assurance? How can engineers develop sufficient confidence to provide assurance about the safety of a new analysis technique or new technology? Can this level of confidence be quantified and measured? This paper has touched on these questions, but has not resolved them.

Engineering classically includes fields such as civil engineering, chemical engineering, and electrical engineering. There are many other disciplines which are part of, or similar to, engineering. I have argued that computer systems engineering (including software engineering and hardware engineering) is a part of engineering, but other disciplines such as synthetic biology or genetic engineering may also fit the epistemological models proposed here. All of these develop and grow objective knowledge to support the design of artefacts (broadly construed), to meet requirements. Medical research too, while usually construed as being scientific, has, in the development of medical therapies, issues of design and assurance which are very similar to those in engineering. Still, the nature of requirements in medicine (for health) is perhaps somewhat less contingent than in engineering. Examining epistemological issues in these subtly contrasting discipline areas may help us to develop a more complete understanding not just of these disciplines, but also of epistemology.

Acknowledgments

NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

References

Anderson, R. (2008). *Security Engineering: A Guide to Building Dependable Distributed Systems*. Wiley, 2nd edition.

Barber, E. H. E., Greenwood, J. N., and Matheson, J. A. L. (1963). Report of the royal commission into the failure of kings bridge. Technical report, Victorian Royal Commission into the Failure of Kings Bridge.

Bell, D. E. and LaPadula, L. J. (1973). Secure computer systems: Mathematical foundations. Technical Report MTR-2547, MITRE.

Boon, M. and Knuuttila, T. (2009). Models as epistemic tools in engineering sciences. In Meijers, A., editor, *Philosophy of Technology and Engineering Sciences*, volume 9 of *Handbook of the Philosophy of Science*, pages 693–726. Elsevier.

Brooks, Jr, . F. P. (1996). The computer scientist as toolsmith II. *Communications of the ACM*, 39(3):61–68.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press.

Clausen, J. and Cantwell, J. (2007). Reasoning with safety factor rules. *Techné: Research in Philosophy and Technology*, 11(1):55–70.

Constant, II, E. W. (1984). Communities and hierarchies: Structure in the practice of science and technology. In Laudan (1984b), pages 27–46.

Constant, II, E. W. (1999). Reliable knowledge and unreliable stuff. *Technology and Culture*, 40(2):324–357.

Davis, M. (2010). Distinguishing architects from engineers: A pilot study in differences between engineers and other technologists. In van de Poel, I. and Goldberg, D. E., editors, *Philosophy and Engineering: An Emerging Agenda*, volume 2 of *Philosophy of Engineering and Technology*, pages 15–30. Springer.

Ferguson, E. S. (1992). *Engineering and the mind's eye*. The MIT Press.

Gainsburg, J., Rodriquez-Lluesma, C., and Bailey, D. E. (2010). A "knowledge profile" of an engineering occupation: temporal patterns in the use of engineering knowledge. *Engineering Studies*, 2(3):197–219.

Helton, J. C. and Oberkampf, W. L. (2004). Alternative representations of epistemic uncertainty. *Reliability Engineering and System Safety*, 85:1–10.

Hoare, C. A. R. (1996). The logic of engineering design. *Microprocessing and Microprogramming*, 41:525–539.

Houkes, W. and Vermaas, P. E. (2009). Produced to use: combining two key intuitions on the nature of artefacts. *Techné: Research in Philosophy and Technology*, 13(2):123–136.

Hughes, T. P. (1976). The science-technology interaction: The case of high-voltage power transmission systems. *Technology and Culture*, 17(4):646–662.

Johnson, A. (2009). *Hitting the Brakes: Engineering Design and the Production of Knowledge*. Duke University Press.

Kroes, P. (2002). Design methodology and the nature of technical artefacts. *Design studies*, 23:287–302.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In Lakatos, I. and Musgrave, A., editors, *Criticism and the Growth of Knowledge*, pages 91–196. Cambridge University Press.

Lakatos, I. (1976). *Proofs and Refutations*. Cambridge University Press.

Laudan, R. (1984a). Cognitive change in technology and science. In Laudan (1984b), pages 83–104.

Laudan, R., editor (1984b). *The Nature of Technological Knowledge*. D. Reidel.

Laymon, R. (1989). Applying idealized scientific theories to engineering. *Synthese*, 81:353–371.

Layton, E. (1971). Mirror-image twins: The communities of science and technology in 19th-century America. *Technology and Culture*, 12(4):562–580.

MacKenzie, D. (2001). *Mechanizing Proof: Computing, Risk, and Trust*. The MIT Press.

Marshall, R. D., Pfrang, E. O., Leyendecker, E. V., Woodward, K. A., Reed, R. P., Kasen, M. B., and Shives, T. R. (1982). Investigation of the kansas city hyatt regency walkways collapse. Technical Report 143, U.S. Dept. of Commerce, National Bureau of Standards.

Mayo, D. G. (1996). *Error and the Growth of Experimental Knowledge*. The University of Chicago Press.

McLean, J. (1985). A comment on the "basic security theorem" of Bell and LaPadula. *Information Processing Letters*, 20:67–70.

Oddie, G. (1981). Verisimilitude reviewed. *The British Journal for the Philosophy of Science*, 32:237–265.

Petroski, H. (1996). *Invention by Design: How Engineers Get from Thought to Thing*. Harvard University Press.

Petroski, H. (2012). *To Forgive Design: Understanding Failure*. The Belknap Press of Harvard University Press.

Pirtle, Z. (2010). How the models of engineering tell the truth. In van de Poel, I. and Goldberg, D. E., editors, *Philosophy and Engineering: An Emerging Agenda*, volume 2 of *Philosophy of Engineering and Technology*, pages 95–108. Springer.

Polanyi, M. (1958). *Personal Knowledge: Towards a Post-Critical Philosophy*. Routledge.

Pols, A. (2010). Transferring responsibility through use plans. In van de Poel, I. and Goldberg, D. E., editors, *Philosophy and Engineering: An Emerging Agenda*, volume 2 of *Philosophy of Engineering and Technology*, pages 189–203. Springer.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. Routledge, 3rd edition. printed 2002.

Popper, K. R. (1963). *Conjectures and Refutations*. Routledge, 2nd edition. printed 2002.

Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach*. Oxford University Press.

Popper, K. R. (1977). The worlds 1, 2 and 3. In Popper, K. R. and Eccles, J. C., editors, *The Self and Its Brain: An Argument for Interactionism*, pages 36–50. Routledge.

Popper, K. R. (1978). Three worlds. The Tanner Lecture on Human Values. [Online at http://tannerlectures.utah.edu/_documents/a-to-z/p/popper80.pdf Last accessed 18 Jan 2014].

Rapp, F. (1981). *Analytical Philosophy of Technology*. D. Reidel.

Rittel, H. (1972). On the planning crisis: Systems analysis of the 'first and second generations'. *Bedriftsøkonomen*, 8:390–396.

Rushby, J. (2013). Mechanized support for assurance case argumentation. In *Proceedings of the 1st International Workshop on Argument for Agreement and Assurance*. Springer.

Staples, M. (2014). Critical rationalism and engineering: ontology. *Synthese*, 191(10):2255–2279.

van de Poel, I. (2010). Philosophy and engineering: Setting the stage. In van de Poel, I. and Goldberg, D. E., editors, *Philosophy and Engineering: An Emerging Agenda*, volume 2 of *Philosophy of Engineering and Technology*, pages 1–11. Springer.

Vermaas, P. E. (2010). Focussing philosophy of engineering: Analyses of technical functions and beyond. In van de Poel, I. and Goldberg, D. E., editors, *Philosophy and Engineering: An Emerging Agenda*, volume 2 of *Philosophy of Engineering and Technology*, pages 61–72. Springer.

Vincenti, W. (1990). *What Engineers Know and How They Know It*. John Hopkins University Press.

Wimsatt, W. C. (2007). False models as means to truer theories. In *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Harvard University Press.