

Science, engineering, and statistics.

by

TP Davis

Henry Ford Technical Fellow for Quality Engineering

Ford Motor Company

(tdavis5@ford.com)

Abstract: Symmetry, parsimony, and unification are important principles that govern the character of physical law. We show how these principles can be applied in engineering to develop a framework that centres on the identification and avoidance of failure modes through design. To support this approach, a definition of reliability, not in terms of probability, but rather based on physics, geometry, and the properties of materials, will be emphasised. We will also show how the nature of the inductive-deductive learning cycle provides the framework for statistical science to be embedded into engineering practice, with particular regard to improving reliability through failure mode avoidance.

1. Introduction

The principles of symmetry, parsimony, and unification are key to scientific progress since they govern the character of physical law, thereby promoting a fundamental understanding of complicated things. In this article, we show how these principles can be applied to the complicated business of automotive engineering and manufacturing, particularly from the standpoint of discovering and avoiding failure modes. Within the appropriate paradigm, and motivated by these principles, failure modes can be seen as a fundamental quantity in engineering, much the same way, perhaps, as electrons are in chemistry and energy and mass are in physics.

Improvement in the understanding of complicated things usually involves iterations between deduction (theory to data) and induction (data to theory). When viewed against a background of variability and uncertainty, which is prevalent in automotive engineering for reasons that will be explained, this iterative sequence can get distorted, and therefore statistical science (the branch of science that deals with understanding variability) needs to be embedded into engineering practice to cope with this. This in turn generates the concept of *statistical engineering* as a treatment for problems in automotive, and other, engineering applications.

Improvement in reliability throughout the service life of vehicles is one of the most difficult technical challenges facing the automotive business. A definition of reliability will be introduced and expanded upon, one that has existed for longer than the accessible literature suggests, and which can be seen to be consistent with the aforementioned scientific principles of symmetry, parsimony, and unification. We will show that these principles will necessarily involve the fundamentals of physics, geometry, and the properties of materials, thereby providing the necessary framework for guiding engineers to take the design decisions necessary to formulate counter-measures to avoid failure modes. This is in contrast to the more "traditional" definitions of reliability that exist, which are based on constructs of probability aimed at measuring the frequency of failure, and are complicated to interpret and measure, and provide no framework for engineers to plan their work.

The ideas presented will be illustrated with examples and case studies taken from current practice within, and from across, the Ford Motor Company.

2. What is engineering?

We begin with this fundamental question. In this author's opinion, the definition offered by the Accreditation Board of Engineering and Technology (ABET) in the United States provides the most succinct description of engineering for the purposes of this article, namely;

“Engineering is the profession in which a knowledge of the mathematical and natural sciences, gained by study, experience, and practice, is applied with judgment to develop ways to utilize, economically, the materials and forces of nature for the benefit of mankind.”

This definition implicitly captures the synthesis between deduction (the laws of the mathematical and natural sciences) with induction (the use of experience, practice and judgment); we will discuss the importance of deduction and induction in engineering more explicitly in Section 4.

A useful interpretation of engineering within the context of science and mathematical analysis, has been provided by Mischke [1]. This diagram is reproduced in Figure 1.

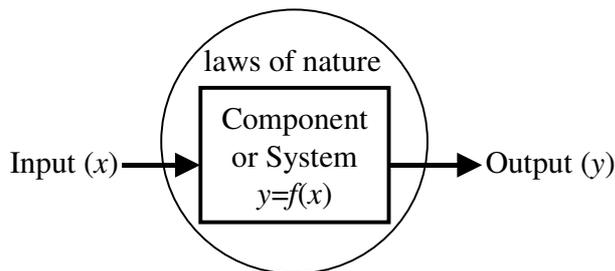


Figure 1: Diagram from Mischke [1], used in explaining engineering within the context of science and analysis (see text). The idea of a *transfer function*, $y=f(x)$, has been added here to Mische's original picture. The transfer function captures the mathematical relationship between the input and the output, and represents the laws of nature as they apply to the case at hand. Of course, the form of the transfer function itself will depend on the actual component or system chosen.

Mischke eloquently explains [1, pages 21-22] that the "name of the game" of engineering is uniquely different from that of science and analysis, but at the same time he explains how engineering requires the use of science and analysis to be effective. This connection is crucial for the development of the ideas we set out in this paper, and so we repeat Mischke's idea here.

He proceeds as follows; if we are given the input, the laws, and the system, and our task is to determine the output, then the skill required is *deduction*, and name of the game is *analysis*. Conversely, if we know the output and need to determine the input, this is still deduction, which we could call reverse analysis. Both of these activities will make use of the transfer function $y=f(x)$, the mathematical representation of the laws of nature as they apply to the case at hand,

although we make the observation that the transfer function may, in practice, be difficult to determine.

Alternatively, if we are given the input and the output, and the system, and our job is to find the laws that govern how the system works, the skill required here is *induction* and the name of the game is *science*.

Finally, if we are given the input and the output, and the laws, and our task is to create a system (perhaps that does not yet exist) that provides the right output for the right input, then the skill required is *synthesis*, and the name of the game is *engineering*.

Mischke's ideas are summarized in Table 1.

| Given | To find | Skill needed | "Name of the game" |
|-----------------------|----------------|---------------------|---------------------------|
| Input, system, laws | Output | Deduction | Analysis |
| Input, system, output | Laws | Induction | Science |
| Input, output, laws | System | Synthesis | Engineering |

Table 1. From Mischke [1], explaining the "name of the game".

Based on the definition of engineering due to ABET, and Mischke's discussion, we can see two important ideas emerging – firstly, that there is a natural connection¹ between science, engineering and analysis (we show later that analysis needs to be supplemented with statistical ideas), and secondly, why the important synthesis of deduction and induction is necessary to solve engineering problems. We use these two ideas to explore how the *principles* of science and analysis (including statistical analysis) can contribute to engineering practice.

3. Some principles of science

A fundamental characteristic of science that distinguishes it from other forms of human enquiry is the requirement that theory has to be confirmed by experiment. It is recognized that progress in theoretical understanding is greatly accelerated through some fundamental principles related to the character of physical laws, among them symmetry, parsimony, and unification, together with the fact that natural laws can be universally expressed in the language of mathematics. We now take a brief look at each of these principles in turn.

¹ The connection between science and engineering can sometimes appear fraught. At the Midlands Air Museum, in Coventry, UK, there is a facsimile on display of a letter, dated June 10, 1940, from the National Academy of Sciences in the US, regarding the potential of the gas-turbine (jet) engine. Part of the facsimile reads "In its present state...the gas turbine could hardly be considered a feasible application to airplanes mainly because of the difficulty in complying with the strict weight requirements imposed by aeronautics. The present internal combustion engine equipment used in airplanes weighs about 1.1 pounds per horsepower, and to approach such a figure with a gas turbine seems beyond the realm of possibility with existing materials". Underneath is a handwritten comment by a local engineer from Coventry; "Good thing I was too stupid to know this" – Frank Whittle.

Symmetry

Symmetry is an important idea that guides the discovery of fundamental laws. Symmetry has an obvious everyday meaning, and the scientific definition is just an extension of this, and is due to Hermann Weyl; symmetry means that after a sequence of operations has been made to a state of nature, certain things look the same as they did before the operations were carried out. In fact, all symmetry laws have a corresponding conservation law related to some quantity, and vice versa, a result due to Emmy Noether. An obvious symmetry law important in engineering is that for a dynamic system of objects moving under the effects of forces, both energy and momentum are conserved – the net amounts of both quantities are unchanged over time. We will use this definition of symmetry shortly, as a way to help define a fundamental quantity related directly to what engineers have to do in creating and developing systems according to the model proposed by Mischke in Figure 1.

Parsimony

Parsimony is the idea of making things as simple as possible (but not simpler), an important feature in science attributed to the mediaeval philosopher William of Occam around 1300 ("one should not increase, beyond what is necessary, the number of entities required to explain anything"). The importance of parsimony has recently been given prominence in the book *Scientific method in practice*, by Gauch [2]. A classic example of parsimony is the replacement of the geocentric model of the universe favoured since the time of Ptolemy, with the heliocentric model re-discovered and published by Nicolaus Copernicus in 1542. It is an interesting point that the heliocentric model was favoured initially because it was simpler to understand, and not because it made more accurate predictions than the geocentric model which, at first, it didn't. The prediction errors in the heliocentric model came about because initially, the planetary orbits were assumed to be circles. Johannes Kepler made attempts to fix these errors using epicycles (the classical "counter-measures" that had been used to fix prediction errors in the geocentric model), but even after these corrections, the remaining errors (in particular with respect to the planet Mars) were still too large to be explained by errors in the data (which had been meticulously collected by Tycho Brahe). Kepler then realised, through careful analysis of Tycho's data, that the orbits were in fact ellipses, and when these were substituted for the assumed circular orbits, the need for epicycles disappeared, leaving a very parsimonious view of the system governing planetary orbits. This parsimony is reflected in Kepler's three laws of planetary motion² (which eventually Isaac Newton was able to reduce to two).

There are many other instances in the history of science that illustrate the idea of parsimony. For example, the simple ratio 9:3:3:1, which determines the proportions of certain hereditary characteristics in living things after two reproductive cycles, was discovered by Mendel in his famous pea experiments. But these ratios were not observed in Mendel's experiments directly – his extensive data yielded estimates close to these numbers, but the set-up of his experiments gave answers that were not quite in these neat integer proportions. It was Mendel's brilliant insight to realize that Mother Nature was parsimonious in the way she constructed her laws, and

² Kepler's three laws of planetary motion are as follows: - 1) orbits are ellipses, 2) each planet sweeps out equal areas in equal times, and 3) the square of the time for a complete orbit is proportional to the cube of the distance from the sun.

he concluded that it was integer proportions that were therefore being uncovered in his data. More detail on this example is given in Gauch [2, pages 288-291].

And, at about the same time as Mendel was growing peas, James Clerk Maxwell was developing the theory of electricity and magnetism. In his four equations describing the electromagnetic field, a constant emerged that was calculated to be very close to the experimentally measured speed of light. Close? Again with brilliant insight, Maxwell realized that it would be unlikely (because it would be un-parsimonious) for Mother Nature to yield two (large) constants in her laws that were close in value to one another, and Maxwell rightly concluded that he had actually derived the theoretical speed of electromagnetic radiation, of which light is a special case (that is, the visible part of the spectrum). This example is given by Mahon [3, pages 107-110], in his recent biography of Maxwell.

By analogy, parsimony in engineering design is also important; one should not, for example, create a system that is over complicated in terms of number of moving parts, and made of unnecessarily expensive materials. In a recent text, Ulrich and Eppinger cite an index of engineering parsimony (due to Boothroyd and Dewhurst) as being proportional to the ratio of the *theoretical* minimum number of parts to the estimated assembly time of the *actual* design [4, page 224]³. Small values of this ratio point to an unparsimonious design.

Interestingly, the principle of parsimony has long been understood in the field of statistical science where the simplest empirical model that explains the data and offers predictions for new observations is favoured over more complicated alternatives; for parsimonious models yield smaller errors in predictions than complicated ones, even though the more sophisticated model fits the data better. Parsimony, then, provides a natural connection between science, engineering, and statistics.

Unification

Finally, the idea of unification (and with it, synthesis) can be seen as providing important structure for understanding the way the world works. This is the idea of combining apparently different concepts into the same theoretical framework, so that more things can be understood with less. Perhaps the most famous example of this idea is Newton's laws of gravitation, which unified terrestrial and celestial mechanics – Newton showed, by studying the orbit of the moon around the earth, that the force that kept the moon in its orbit was exactly the same as the force that made things (apples, for example) fall to the ground on earth⁴. Other examples include combining light, electricity & magnetism into one set of laws of electromagnetism by Maxwell, combining the laws of energy and mass into the generally theory of relativity by Einstein, and

³ Only parts satisfying at least one of the following conditions must be theoretically separate: - i) does the part need to move relative to the rest of the assembly?, ii) must the part be made of a different material from the rest of the assembly?, and iii) does the part have to be separated from the assembly for access, replacement or repair?

⁴ It was as a consequence of the discovery of the law of gravitation that Newton was able to reduce Keplers three laws of planetary motion down to two, thus enabling him to understand more with less – it turns out that elliptical orbits are a consequence of the force due to gravity pointing toward the center of the sun (which is the law that produces equal areas in equal times), together with the force varying inversely as the square of the distance from the sun (which is the law that produces orbital periods that are a $3/2$ power of the distance from the sun).

the development of quantum theory, which provides a single treatment for waves and particles within the laws of quantum mechanics.

We will use the idea of unification (to understand more with less), together with symmetry (and its implied conservation law of fundamental quantities), and parsimony (for simplification of complicated phenomena) to enhance the understanding of the framework for engineering given by Mischke [1]. In particular we will introduce a treatment for reliability, which requires a different approach than the one traditionally used based on (a frequency interpretation of) probability.

4. Deduction and induction

The ideas discussed above begin to hint at the interplay between deduction and induction in the scientific method. Deduction argues from a given theory's general principles to a specific case of expected data. It moves from the mind to the real world, from theory to data. For example, consider the following quote from [5] regarding vehicle dynamics-

"We view dynamics as a deductive discipline, knowledge of which enables one to describe in quantitative and qualitative terms how mechanical systems move when acted upon by given forces, or to determine what forces must be applied to a system in order to cause it to move in a specified manner."

Induction on the other hand argues in the opposite direction, from actual observed data, to an inferred model. It moves from the real world to the mind, from data to theory. This is the domain where statistical science can be most effective because data are collected either from direct observation of the system or phenomena under study, or through directed experiments with the system itself⁵, allowing approximating mathematical models to be developed which can be used to describe (at least approximately) how these systems might work, thus enabling predictions of future observations to be made. For an extensive discussion of deduction and induction within the context of the scientific method, see [2].

The interplay between deduction and induction provides part of the synthesis required for engineering to make use of the combination of analysis with science. Indeed, this interplay also characterizes the interplay between theory and experiment in the pure scientific sense. While engineering problems are not concerned with the discovery of new fundamental laws, engineers do need to understand how the known laws combine and apply in their particular case, which in itself can be very complex. George Box ([6], and in many subsequent publications) has eloquently discussed the interplay between deduction and induction from the standpoint of using statistical science as a catalyst for discovery. Figure 2 is based on Box's diagram in [6], but with the significant modification that the lines representing theory and data *converge* to a solution

⁵ Kepler's laws of planetary motion were discovered as a consequence of the analysis of direct observations of the positions of the planets in the sky. Newton's discoveries regarding the decomposition of light came about through him experimenting with prisms in a darkened room. These directed experiments lead Newton to formulate a fundamental and new theory of optics, whereby white light was shown to be made up of a combination of coloured light, and was not "pure" as had previously been thought.

(they are parallel in Box's original model). The solution, of course, is one that "utilizes, economically, the materials and forces of nature for the benefit of mankind".

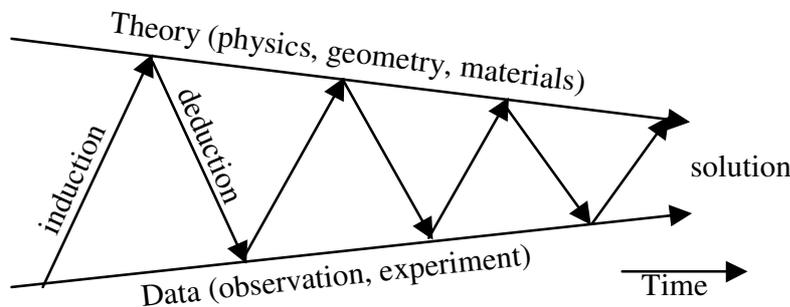


Figure 2: a model for learning through iterating between theory and data using induction and deduction (after Box [6], but here with the lines shown converging)

There is a natural connection between using deductive and inductive thinking to learn about new phenomena, and in so doing making inferences about the way the world works – sometimes one leads, sometimes the other, but it is difficult to perceive one being useful without the other. Indeed, in this authors experience, engineering investigations often get bogged down because the team get stuck either in the deductive or the inductive mode, and a "nudge" to get them to consider the other state is often enough to re-energize and accelerate the investigation toward a solution. For example, we do not need to run factorial experiments to derive an empirical representation of the way the system works if we can deduce the answer from already known theory. By the same token, if we *can't* deduce a solution to a problem, or deductions provide predictions that are logically inconsistent, we should be prepared to conduct experiments to generate ideas for new hypotheses and theories as to how the laws of nature might apply in the case under investigation.

In the automotive engineering world in particular, the iterative sequence implied by Figure 2 is relatively immediate (so that experiments and observation can be made quite quickly), but executing this process and interpreting the results is made more complicated by the presence of variability. As we have just remarked, in automotive engineering, this variability is caused by production conditions (mainly the *rate*), hardware complexity (including variety of product combinations), and the large numbers of units in the field exposed to uncontrolled (and largely unknown) usage profiles and demand cycles. We will continue the discussion of variability later, under the general heading of noise factors, but for now we recognise that variability distorts predictions from both deduction and induction; and to deal with this, we therefore have to appeal to the branch of science that deals with variation - statistical science.

Before proceeding however, we note that the application of statistical science to any field of human enquiry is highly contextual. The iterative cycle of learning illustrated in Figure 2 is relatively immediate in automotive engineering, unlike, say, agriculture, where it takes at least a year to complete an experiment (the inverse square law of gravity at work here!). This particular contrast between agriculture and engineering is important, because many statistical methods were developed in the early part of the 20th century specifically with agricultural applications in mind, but we should not necessarily expect that these methods could be adopted within the

context of engineering without some modification⁶. The importance of contextual considerations with regard to the application of statistical science has recently been re-emphasised by Sir David Cox, in brief remarks commenting on an article by Hahn [7, page 298].

5. What is statistics?

Having discussed the importance of statistical science in Sections 3 and 4, we now try to define it.

We begin by noting that statistics, as part of *scientific* activity, is not to be confused with probability, which is primarily a *mathematical* concept. When the probability under consideration has a frequency interpretation, Gauch [2] refers to it (the probability) as a deductive measure of uncertainty. This can be seen for example, when physicists use probability to measure variations in positions and velocities of particles at the atomic and sub-atomic level; they then rely on averaging over a large number of particles, to check whether their experimental results match theory. An early example of this is the determination, by Maxwell, of the velocities of molecules in the development of the kinetic theory of gases. Maxwell derived a probability model to explain the distribution of velocities of a large number of molecules in a closed container. The probability distribution so developed now bears his name.

An alternative use of probability is as a measure of a degree of belief about the state of nature. A simple example (given in [2]) to illustrate this point is as follows. Consider two problems:

Problem 1: Given that a coin is a fair coin (the hypothesis, H), what is the probability that 100 tosses of the coin will produce 45 heads and 55 tails (the data, D)?

In shorthand, we could write this problem as Probability(D given H), or $\Pr(D|H)$. To answer this question we appeal, without any need to understand the context in which the question is asked, to the mathematics of the binomial distribution, and deduce, through calculation, that the answer is 0.0485. This calculation, assuming the coin tosses are independent and not designed to favour one outcome over the other, is exact to four decimal places. The result does not have to be confirmed by experiment. This is a probability problem that can be solved completely with mathematical deduction. The probability so calculated measures the *frequency* with which the result (exactly 45 heads) would occur over a large number of repeated trials with the same coin.

Problem 2: Given that 100 tosses of a coin produces 45 heads and 55 tails (D), what is the probability that the coin is fair (H)?

We might write this, in a similar way to above, as $\Pr(H|D)$. Note that the roles of H and D in this probability statement are reversed as compared to problem 1. Solving problem 2 requires much more than determining this probability through mathematical calculations alone; We now have to understand the context within which we are asking the question - for example, knowledge of the way coins are made, and how they are flipped will enable us to define what we might mean by

⁶ An obvious example is the development of response surface methods to reflect the sequential nature of experimentation in industrial applications, compared to the "single-shot" experiments often conducted in agricultural studies. See [6] for a discussion.

fair or unfair, what the likelihood of finding such a coin is, and the chances of it coming down heads, and so-on. This is a probability problem that can only be solved with some further enquiries and possibly more experiments⁷. The probability so derived measures a *belief* about the state of nature regarding the coin that has been flipped. Unlike problem 1, where the probability statement concerned the data, there is no frequency interpretation here at all, because we want to know something about the *particular* coin that has been flipped and no other, based on the data at hand. The distinction between using probability to measure either frequency or belief has been well explained in a recent book, *Probability theory – the logic of science*, by the physicist ET Jaynes [8]; indeed, as part of his discussion of these two problems, he advocates that frequencies should not be called probabilities at all, to minimize the confusion in interpretation.

We can perhaps loosely relate these two probability problems to the induction-deduction interplay of Figure 2. Deduction can be represented as $\Pr(D|H)$, which is problem 1, while induction is $\Pr(H|D)$, problem 2.

It is, then, of fundamental importance to recognize that $\Pr(D|H)$ is not the same as $\Pr(H|D)$, the nature and context of these probabilities being completely different (as, indeed, are induction and deduction different), although this point is often not well appreciated. For example, many courses in statistics given to engineering undergraduates and to engineers in industry spend a lot of effort teaching statistical significance tests. These tests generate so-called "p-values", which actually provide a measure $\Pr(D|H)$. But in many cases the engineering investigator really wants to make a statement about $\Pr(H|D)$, and the p-values (and associated levels of significance) are of no help here. Never-the-less, these p-values and significance levels are in abundance in much statistical literature and training material. The excellent article by Carver [9] has more details on this subtle point.

The point may be subtle, but it is certainly not academic, as some life-examples will illustrate. Suppose the data is that the patient has spots (S), and the hypothesis is that the patient has measles (M). Obviously, you do not need to be a doctor to realize that $\Pr(S|M) \neq \Pr(M|S)$. The first probability is close to 1. If you have measles you will almost certainly have spots. The second probability is, say, about 0.1, because if you have spots, measles is only one of a number of other diseases you may have that could bring you out in spots (we will return briefly to the medical field later). Or, suppose that the data is the evidence against the suspect (E), and the hypothesis is that the suspect is guilty (G). Obviously, you do not need to be a lawyer to realize that $\Pr(E|G) \neq \Pr(G|E)$; but miscarriages of justice are documented that have arisen from the jurors, whose job is to evaluate $\Pr(G|E)$, mistakenly evaluating $\Pr(E|G)$, perhaps as a consequence of a skilful prosecution lawyer, or misleading arguments from an "expert" witness⁸.

This dilemma also applies to engineering problems, as the following actual example, related to product reliability, will show (this example provides a foretaste of probability issues surrounding

⁷ Most of us would use our experience and judgment to suggest that finding an unfair coin that would still produce heads and tails in near equal quantities in an experiment of 100 tosses would be extremely unlikely, and assign a probability close to 1, very different from 0.0485.

⁸ A recent example, in the UK, is the case of Sally Clark, herself a lawyer, who was convicted of murdering two of her infant children based on incorrect arguments of probability, including a mix-up of these conditional probabilities. For an informative account of the probability arguments in the Clark case, see the article by Helen Joyce [10].

reliability which we address in Section 7). The production of a well-established product was extended to a new facility, intended to be identical to the original, to meet increased market demand. During the final reliability testing of the product assembled in the new facility, a part suffered a structural failure after only about one third of the total duty cycle of the test. It was known that this failure had never been observed on similar testing of the product from the original facility, and furthermore, there was no evidence of these failures being seen in the field. The question to answer then, was – does the product from the new facility have a reliability problem, or is this event simply a chance manifestation of a rare failure, equally applicable to both production sources?

An accelerated bench test was developed which reproduced the failure mode on the broken component. Data from samples of parts from both facilities were tested, and parts from the new facility were seen to be somewhat inferior (that is, they failed earlier) to the performance of parts from the original facility. This, along with the original durability testing results, is the data, D. The hypothesis is that the two manufacturing facilities produce parts with the same reliability in the field, an event which we will denote R. It might be tempting to perform a statistical significance test on the bench data to make a decision regarding the reliability of units from the new facility. Such a significance test would be evaluating $\Pr(D|R)$, but by definition, we cannot condition on R because this is what we want to find out! In fact for the data, D, in this example, such a test yields a p-value of 0.15. We might be tempted to infer that this suggests no significant difference between the two manufacturing facilities⁹, and proceed to ship product from the new facility, since we have had no field failures from the original facility.

But, by making parallels with the previous life examples, what if we recognise that the question we are really trying to answer has the roles of R and D reversed, i.e. we want to know $\Pr(R|D)$? This motivates us to ask questions about whether, for example, the two manufacturing facilities really are nominally identical. In fact, further investigation found out that they were not; some processes that were automated in the original facility were done by hand in the new, and so-on. Hence, it was possible to document differences in the two production facilities that might explain, through deductions based on technical know-how, the inferior results from the new facility. The problem now becomes one of figuring out how to implement counter-measures in the new facility, for example by modifying the manufacturing process, so that the observed failures can be eliminated (i.e. we now recognise that the second facility was different to the first to begin with, and so now knowingly accommodate these differences in our inferences so that the counter-measures necessary to avoid the failures from the second facility can be planned).

This is not to say that there is no place in engineering for significance tests and p-value calculations as taught in statistics courses. As we have seen, deductive logic relies entirely on what is a-priori known, and these are the hypotheses upon which we can condition to make judgements about data using such tests. For example, we know that Newton's 2nd law will apply when calculating the acceleration of an object of known mass subject to a known force. Knowing this may be useful in solving some problems directly, for example specifying the size of a brake rotor to stop a truck of known mass travelling at known speed in a given distance. We might make such calculations, and then to confirm our analysis, build and test a proto-type. We would collect data on stopping distances. We want to know if these data are consistent with our

⁹ p-values of 0.1, or sometimes 0.05, are often used to highlight a "significant" result.

analysis, i.e. we want to know something about $\Pr(D|H)$. So a p-value from a significance test is useful here; if our p-value is low (say <0.1), we might conclude that our analysis in designing the brake was faulty, and therefore must check our work. We certainly would not reject the hypothesis, for we know Newton's 2nd law is true, whatever our data¹⁰!

In summary, inferences that are targeted at making statements about $\Pr(D|H)$ can be well served by the mathematical deductions of statistical significance testing that result in the familiar p-value measurements of probability, interpreted as a frequency. Inferences regarding $\Pr(H|D)$ on the other hand, where the probability now refers to the belief about the state of nature under consideration (i.e. the hypothesis, H), requires a different approach than just the application of mathematical calculations to derive the probabilities, and will necessarily combine data from the current experiments with the stated level of technical knowledge of the problem under investigation.¹¹

We will therefore define statistics as

"...the science of making inferences through inductive logic and reasoning in the face of uncertainty".

and begin to explore how this view of statistics can help with engineering.

6. Statistical Engineering

In the presence of variability, it is problems that require the use of inductive logic (extracting new, empirical knowledge about something that cannot be easily deduced from what is already known) that require careful application of statistical science, because we are trying to say something about $\Pr(H|D)$, where the data might itself be subject to some uncertainty. The mathematical calculations that lead, for example, to significance levels and p-value's, don't help with this problem, as we have seen. What is needed is not so much methods aimed at testing hypotheses, but methods directed toward actually *generating* them. Inevitably, these hypotheses will come about through looking carefully at the data, and in conjunction with subject matter

¹⁰ Another area where p-values find utility is in assessing the adequacy of both theoretical and empirical models with reference to the experimental data through derived quantities such as residuals (the differences between the actual observations and predictions from the model). A famous example in science concerns the discovery that experimental data, across a range of temperatures, slightly but consistently exceeded predictions from the (assumed to be true) Wein's law, which models the intensity of sealed-cavity radiation as a function of wavelength. Max Planck showed that these discrepancies between model and experiment could only be explained by assuming that energy came in discrete packets (quanta), and quantum mechanics was born. For an account, see [11, Ch. 1].

¹¹ Some readers may well recognize this distinction as the difference between the frequentist and Bayesian schools of statistical inference, both developed in the first part of the 20th century, primarily by RA Fisher, who was a frequentist, and H Jeffries, who was a Bayesian. In [8], Jaynes explains how the contextual differences of their respective sciences led to these two approaches. Fisher was a pre-DNA geneticist and biologist, who was trying to solve frequency-related problems without much prior knowledge available to guide experiments, while Jeffries, who was a physicist, had 200 years of Newtonian mechanics and electromagnetic theory to draw on in formulating his alternative approach to inference. Jaynes offers some anecdotal evidence based on the personalities involved to explain why the frequentist approach came to dominate the Bayesian one in the literature, and subsequently in much scientific practice.

knowledge already established, postulating models that provide a possible explanation of the structure therein, so that these models can be used to make a cautious prediction on what the next experiment might yield. As we have previously suggested, data can be collected either through informed observation of the mechanism under study, or by directed experimentation with it. These experiments are carefully designed using statistical principles to make the system behave differently so that it can be better understood. Statistically designed experiments depart from the perhaps conventionally held wisdom that experiments have to hold all variables (or factors) constant, and vary just one at a time to find out how things work. Statistically designed experiments vary all factors simultaneously, in order to uncover otherwise hidden features of the mechanism under investigation, particularly with regard to how the factors may interact (that is, the effect of one factor may well depend on the state of another).

The requirement to achieve synthesis between deduction and induction in the presence of variability, provides an invitation for statistical science to be embedded into engineering practice. To make advancement in knowledge, answering the question $\Pr(\text{HID})$ requires a combination of inductive reasoning from empirical evidence, coupled with deductions based on known facts and theory – in other words, the scientific method. Statistical science provides the framework for inductive inferences, and this approach has been called *statistical engineering* in [12]. To quote directly,

“The combination of engineering science (the study of physics, [geometry] and materials) and statistical science (the empirical modelling of variability) is necessary to achieve what is demanded from us by our customers - a consistent level of superlative performance.”

Just what *is* demanded of us from our customers? A recent survey conducted by JD Power in the United States [13], lists several vehicle attributes that customers consider to be the most important purchase decision factors when buying a new car. The top few items in this list, are in order of importance; Drive, handling, & vehicle performance; Comfort; Styling & design of the exterior; Safety; Long-term reliability; Physical dimensions (e.g., seating room, cargo capacity, height); and fuel efficiency.

Note the presence of long-term reliability in this list. In the same study, JD Power show that with four faults or more, the chances of a customer remaining loyal to their current brand for the next purchase is reduced by half, from 38% down to 19%, so it is clear that reliability is an important attribute as far as customers are concerned, and most readers will be able to relate to this directly. We therefore now turn our attention to the subject of reliability, and make the case for the need to change to an alternative definition for the purposes of ensuring that we improve the reliability of product at the only place that matters – in the field, a highly variable place in the automotive world. We will then construct a framework to enable us to engineer for it, allowing for the presence of this variability. We note that, in practice, engineers are required to make inferences about reliability in the field based on test data collected on the bench on the laboratory, and not the other way round (as we saw in the engineering example of Section 5), and therefore it is natural that statistics, as we have defined it here, will be useful, not as a replacement for engineering knowledge, but as a catalyst for it [6].

7. What is reliability?

The definition of reliability that appears in many textbooks on the subject is close to the following:-

“reliability is the probability that the system performs its intended function for a specified period of time under a specified set of operating conditions”.

It usually appears in the first few pages, for example, see [14], [15] and [16]. This is a definition of reliability in terms of a survival probability. Since in practice this definition refers to the state of more than one unit in the field, the probability interpretation here is one of frequency rather than a degree of belief, and therefore using the arguments in Gauch [2] and Jaynes [8], is a deductive measure of uncertainty about reliability. It has a long history of application in the nuclear and aerospace engineering fields. But we note that in these industries, unlike automotive which is our primary concern here, there are relatively small number of units in the field, performing under specified and tightly controlled operating conditions, with closed-loop feedback from *every unit* in the field, over their *entire* service life. As we have mentioned, the context of automotive engineering is very different; in addition to production rates, product complexity, and variable operating conditions in the field, in the automotive world there is the added problem of poor field data of actual performance in terms of failures, which does not extend to every unit over the entire life¹².

But this definition of reliability (i.e. via a survival probability) cannot be operationalized in the engineering design process without being able to make predictions of it, and to subsequently be able to confirm it by experiment. Even making an assessment of the current state of reliability in the field (as a frequency, or rate of failures) requires having access to either all of the field data, or at least a representative sub-set. But this is denied to automotive engineers – since, for example, all vehicles eventually leave the warranty period and so are lost to follow-up, and these are the higher mileage and older vehicles that arguably, given the JD Power survey [13], we should be most interested in. As an added problem, time is often not the appropriate life variable, mileage being more relevant, but miles accumulate at different rates for different customers. But even if we *could* measure the reliability in the field, we would then have to make predictions of it from laboratory and prototype testing during product development. This is never possible in practice because estimating probabilities empirically requires large sample sizes (perhaps not a problem for particle physicists, but certainly for automotive engineers!), and so, with the probabilistic definition of reliability guiding our actions we are left in the unsatisfactory position of not being able to confirm theory with experiment, either on the bench or in the field. In fact, this problem has also been identified in aerospace; see for example Richard Feynman's contribution to the Roger's report into the Challenger space shuttle disaster of 1986 [17]. One of Feynman's crucial observations was that trying to measure failure rates and frequencies with probabilities that could not be directly observed, creates a tendency to assume that the reliability (in the field) is better than it actually is, a "travelling hopefully" outcome to the dilemma of not being able to measure what we are trying to make inferences about.

¹² Some data on field failures is of course captured through warranty and customer surveys, but these only provide a partial glimpse of the actual number of failure in the field, and are not comparable to the data capture mechanisms available in aerospace and nuclear engineering. These data sources do, however, provide useful information on the types of failure modes present in the field.

The probabilistic definition of reliability also lacks parsimony, reflected through the fact that it has to qualify the probability as being conditional on a specified period of time and a specified set of operating conditions. This lack of parsimony can be illustrated by a mathematical development of the probability implied by the definition, much in the same way as we did for coin-flipping using the binomial distribution, although here of course the mathematics needed is much more complicated. The reliability as defined represents the probability that the survival time (which we will denote in upper case by T , although we note this might also refer to mileage or usage cycles) exceeds a specified period of time (which we will denote in lower case by t), under specified operating conditions (which we will denote by N). We can write this as Probability(T exceeds t , for given N) or $\Pr(T>t|N)=R_{TN}(t|N)$. But because in the field we cannot condition on N , (i.e. we can't find customers who only use their vehicles according to our specific operating condition) we can only measure at best the average, or marginal, reliability, which is Probability(T exceeds t , averaged over all N) = $\Pr(T>t) = R_T(t)$. Without going into details here (but see [18]), these quantities are related through the following equation of probability densities, $f(\cdot)$:-

$$R_T(t) = \int_t^{\infty} \int_{n \in N} f_{T,N}(t,n) dn dt = \int_t^{\infty} \int_{n \in N} f_{TN}(t|n) f_N(n) dn dt = \int_{n \in N} R_{TN}(t|n) f_N(n) dn \quad (1)$$

Solving this integral needs determination of quantities that are impossible to attain, specifically $R_{TN}(t|N)$ which is the survival probability at every conceivable noise condition, and $f_N(n)$, which is the probability distribution of all these noise conditions in the field. These quantities are unknown and unknowable thus making the traditional definition of reliability unworkable in making inferences about $R_T(t)$, which is essentially the point emphasized by Feynman. The equation (1) also illustrates the idea that reliability in the field, $R_T(t)$ is not the same thing as the reliability demonstrated on a bench under a specific noise condition, $R_{TN}(tn)$, a point that often gets overlooked in translating inferences from bench testing to failure rates in the field.

As we have seen from the JD Power study [13], reliability is one of several vehicle attributes of importance to customers. Engineering descriptions of the attributes in [13] include, for example, Package (measured by m^3), Ergonomics (N, m, s, etc), Vehicle dynamics (N, Nm, etc), Performance (ms^{-2}), Fuel economy (l/100km), Emissions (gm^{-1}), and so-on. All of these attributes are measured with appropriate scientific units (given in parentheses) that are related to the physics, geometry, and properties of materials¹³ that govern the attribute. A natural question to ask, therefore, motivated by the scientific principle of unification, is why should reliability be different? Probability is a unitless quantity, unrelated to physics, geometry, and material properties, and so under this definition, there is a lack of unification in the way the attributes are described and measured.

We argue that this lack of unity in the attributes presents us with a problem (but also an opportunity) that is common in science – to find a way of unifying different things (attributes in

¹³ Geometry tells us where we are, physics, through mechanics, tells us where we are going, and the materials we choose will dictate whether or not we get there.

our case) so they can be seen as different aspects of the same whole, enabling us to understand more with less. To do this, we have to define reliability in such a way that we can measure it with physical units, and that, at the same time, have it become more parsimonious than the traditional probabilistic definition.

A very simple (parsimonious) definition of reliability, defined by Don Clausing in the 1970's, but only recently published in the accessible literature [19], is as follows:-

“reliability is failure mode avoidance”

Clausing defines a failure as any customer perceived deviation from the ideal condition and operationalizes this by defining two types of failure mode and two causes for these failure modes. The two types of failure mode are *a*) hard failures (something breaks, or completely ceases to function) and *b*) soft failures (the item still works, but functional performance degrades to a point where a customer will complain)¹⁴. The two root causes of these failure modes are, on the one hand, mistakes (that is, failure to take an action or counter-measure, that is known a-priori to avoid a failure mode) and on the other hand lack of robustness (sensitivity to the variability discussed in Section 4). So a reliable design is one that is as robust and as mistake free as possible. We discuss robustness in more detail in Section 9.

We will show shortly how to measure our ability to avoid failure modes, through a concept called the *distance from the failure mode*. The measurement of these distances will involve physics, geometry and the properties of materials, and in this sense will unify reliability with the other attributes.

Before proceeding however, we offer one other justification of this definition of reliability, through analogy. The context of reliability in automotive engineering is very close to the context of survival in medical studies, where measurements of probabilities are generally not used to measure consequences of proposed actions (surgery, drugs, lifestyle & dietary advice) to eliminate illness and disease and promote good health (failure mode avoidance!). Practitioners in the medical field use measures of *relative risk*, often determined in clinical trials, because *absolute* risk is impossible to measure and quantify, and clinicians recognize that extrapolating from a trial to the larger population can be foolhardy. The idea is to demonstrate an improvement in survival (elimination, or delayed onset, of disease) compared to some prior baseline level. A common model to effect such an analysis is the so-called proportional hazards model¹⁵, where inferences on different survival rates can be made without knowledge of the absolute underlying rate. To illustrate the connection between medical survival and engineering reliability, this model can also be used to solve complicated engineering reliability problems. For a contemporary account of an engineering application of the proportional hazards model see [20].

¹⁴ The examples in Section 11.2 provide illustrations of these two types of failure mode.

¹⁵ Connections between the medical and engineering fields are not as far fetched as might at first appear. Many engineers first meet the hazard function under the guise of the "bathtub curve", which is often used to explain the phases of burn-in, useful life and wear-out in engineering applications [e.g. see 14, page 29]. The origin of the bathtub curve is simply the graph explaining the corresponding phases of human survival (infant mortality, healthy existence, and old age), and the technical name for this failure rate curve is the hazard function. We are not going to discuss the hazard function in this paper, but some details related to engineering applications can be found in [21].

Some comparisons between nuclear, aerospace, and automotive engineering with the medical field, against some reliability criteria, are given in Table 2. It is often thought that automotive engineering reliability can be better understood by trying to replicate practice in the nuclear and aerospace industries (which is one reason, perhaps, why the probabilistic definition pervades), but the analysis of Table 2 would suggest differently, that the medical field is closer in context to the problems faced by automotive engineers.

| Reliability Criteria | Nuclear | Aerospace | Automotive | Medical |
|-------------------------------|----------------|------------------|--|--------------------------------|
| Units in the field | 00 | 000 | 0000000 | 0000000 |
| Quality of field records | | | | |
| Failed units | Excellent | Excellent | Reasonable | Good |
| Unfailed units | Excellent | Excellent | Poor | Poor |
| Units lost to follow up | No | No | Yes | Yes |
| Noise space | Simple | Simple | Complicated | Complicated |
| Life variable | Operating Time | Operating Time | Calendar time Mileage Usage cycles | Calendar time Exposure time |
| Competing risks ¹⁶ | No | No | Yes | Yes |
| Scientific context | Deductive | Deductive | Inductive | Inductive |
| Key reliability measure | Probability | Probability | Distance from the failure mode | Relative Risk |

Table 2: comparisons of automotive reliability with the nuclear, aerospace and medical fields. The contextual similarities between the automotive and medical fields are striking.

Note that directionally the new definition of reliability is consistent with the probabilistic definition in the sense that eliminating failure modes will increase the survival probability, but with Clausen's definition, the focus is on directly addressing the actual failure modes themselves (the *how* and the *why* of things failing), rather than the consequences of the failures measured as a rate or probability. This idea is much more fundamental with respect to what engineering is all about (which ultimately leads to making a single decision – whether or not to release the proposed design for production), and we will illustrate this with several examples in Section 11.

8. Failure mode avoidance and symmetry

We now turn to the idea of symmetry and its implied conservation law, to propose a treatment for dealing with the failure modes to be found in an engineering component, and we use this symmetry law to argue that the only uncertainty is *when* the failure modes will be found, not *if*.

¹⁶ Competing risks refers to masking between two or more failure modes. For example in medical studies, a clinical trial to test a new drug for, say, the relief of hay-fever, may be hampered by the fact that some patients will fall ill for other unrelated illnesses. In automotive applications, competing risks can occur when robustness problems mask those due to mistakes, and vice versa. This latter point can cause many problems in production, particularly if failure modes are discovered very late in development. Competing risk problems in nuclear and aerospace engineering are rare, because robustness problems are not as prevalent.

The process of developing an engineering design and putting it into mass production can be characterized as follows (for example see [4] and [22]);

Concept & Design → Fabrication & Testing → Production Preparation & Manufacture (2)

From the point of view of the engineer, the operations of fabrication & testing, and production preparation & manufacture, should leave the design in an unchanged state from the original intention captured on the engineering drawing. Therefore, engineering can be viewed as if being governed by a symmetry law in the strict scientific sense of the word (things are the same after a series of operations as they were before). Failure modes potentially disrupt this symmetry, because once fabrication & testing and thereafter production preparation & manufacture are executed, the only thing that will force an engineer to make a change to his or her design is the discovery of a failure mode that requires a counter-measure. But a counter measure cannot be developed until the failure mode has been discovered. The earlier the failure mode can be discovered, the earlier the counter measure can be developed. Finding failure modes early is one of the biggest challenges in automotive engineering, because late discovery of failure modes leads to late engineering changes to accommodate the counter-measure. This can translate into inflated design costs (because freedom to adopt the optimal counter measure is severely restricted), poor product launches (because changes to the engineering have to be processed at the same time as getting ready for mass production), excessive warranty (because some failure modes inevitably escape into the field), ultimately leading to low customer satisfaction, and significantly reduced owner loyalty, as we saw in the JD Power study [13].

Symmetry in engineering, in the sense described, can only be preserved if we find and eliminate failure modes during concept and design, rather than the fabrication & testing or the production & manufacture phases. From a practical standpoint, this symmetry is managed (and reflected) in the “V” model (used by most automotive engineering enterprises), which we discuss later (in Section 11.2).

Therefore, a failure mode can be thought of as anything that disrupts this symmetry¹⁷. The conservation law implied relates to the number of failure modes that have to be found. This number, we conjecture, is constant (much the same way that the number of electrons involved in a chemical reaction is constant, as is the amount of energy required to do a prescribed amount of work). The only unknown is when we choose to find these failure modes so that they can be counted and eliminated.

The act of engineering implied by the sequence of activities in (2) therefore embodies three steps:-

¹⁷ This definition of a failure mode is very profound. It not only encompasses the usual definitions of technical failure modes as proposed by Clausen, but also widens the definition to include anything that drives a change to the design late in the engineering process, including mistakes in definition and planning that are only perhaps realized mid-way through a project, and mistakes resulting from poor characterization of the design, which might lead to late design changes aimed at reducing cost or weight, for example. This broader definition of failure modes will be discussed in a forthcoming paper [23].

- Step 1: create the failure mode (unintentional but inevitable)¹⁸
- Step 2: find the failure mode
- Step 3: take an effective counter-measure to avoid the failure mode

Step 1 (almost) always happens during Concept & Design; Steps 2 and 3 happen any time after that (sometimes customers get to execute Step 2). So what does it take to find and eliminate failures early? Mother Nature does not change the laws of physics, rules of geometry, and properties of materials during fabrication & testing and production preparation & manufacture, and so we should be able to use this fact to help find the failure modes (Step 2) during the Concept & Design phase, so that the counter-measure (Step 3) can be taken prior to Fabrication & Testing, therefore preserving symmetry – that is, the design eventually looks the same in the hands of the customer as it did in concept & design phase, despite transitioning through the operations of fabrication, testing, production preparation and manufacture.

Note that this insight provides another compelling reason for the failure mode avoidance definition of reliability to take precedence over the probabilistic alternative, since it can immediately contribute to achieving symmetry in engineering, by focussing attention directly on the failure mode itself (the fundamental quantity), rather than the frequency with which it might occur.

Recall that there are two root causes of failures – mistakes and lack of robustness. Avoiding failure modes due to mistakes is primarily a matter of vigilance, because the occurrence of a mistake is the entropic state in a business as complicated as automotive engineering. Our natural ability to be vigilant is enhanced by, among other things, establishing operational definitions¹⁹, conducting Failure Modes and Effects Analysis (FMEA) to do "thought experiments" on likely causes of failure, avoiding deviations from established design standards, conducting design reviews with peer group experts, maximizing commonality and re-usability of parts so that changes can be minimized, and so on. Usually these mistake avoidance activities are collectively bundled into a standard operating procedure for the business, perhaps labelled as Total Quality Management (TQM), Quality Operating System (QOS), or some other name.

With regard to robustness failures, there are two ways to find such failure modes early in product development. One way is testing. Testing requires hardware and therefore to do it early it should be focussed as much as possible on the components, rather than the integrated systems or the vehicle. Additionally, component testing is more flexible (e.g. larger sample sizes are possible, and a wider range of stresses can be included); and, in the final analysis, only components fail, not systems or vehicles (to appreciate this, recognise that failures might be *discovered* during system or vehicle tests, but counter-measures always require changes to *components*). In particular, the ability to find robustness failure modes during component testing is determined primarily by the ability to apply the stresses (which we will shortly refer to as noises) that the component will see when it is integrated into the system and finally the vehicle. This is really the

¹⁸ Some failure modes will be intrinsic to the basic function of the design; others will depend on the system or component chosen to deliver the function. The only open question is how many failure modes are there? This question will be addressed in [23].

¹⁹ WE Deming, the well-known quality management leader, proffered the following - "An operational definition interprets – it is not open to interpretation".

essence of systems engineering – an appreciation of what a component has to do when it is performing in a larger system.

The second way to find failure modes is through analysis. By its very nature, analysis can be executed earlier than testing. As just mentioned, the well-known technique of FMEA can be used to begin the process of finding failures in this way. Analysis will also require the use of transfer functions, the mathematical representations that relate the output of the component to certain values of the inputs (see Figure 1). The transfer functions are dictated by the laws of nature, and will involve the physics, geometry, and the properties of materials of the design. Constructing the transfer function will rely on our ability either to deduce it (through knowledge of how the laws of nature apply in our case), or on our ability to estimate it empirically through experimental induction.

We will measure failure mode avoidance by counting mistakes and, in the case of robustness, measuring the distance from the failure mode. Our primary interest for the remainder of this article will be in failure modes caused by lack of robustness, because it is the development of counter-measures for these types of failures that require technical analysis through the use of transfer functions and the use of statistical science as a treatment for uncertainty caused by variability. Avoidance of mistakes is primarily a matter of vigilance.

9. Failure modes due to lack of robustness

Robustness is the ability of a component or system to function in the presence of noises (this definition is due to Taguchi [24, 25]). Function involves basic physics, because mechanical engineering is about making things move, or stopping things from moving, together with the associated geometry and the properties of materials. A word is in order on noise factors, since it is these that cause the robustness failure modes. Noise factors are, collectively, the sources of variability we referred to in Section 7. There are basically two types of noise factor; those noises that affect a design's *capacity* to function, and those noises that determine the *demand* against which the design must function.

The capacity noises (called *inner* noises in [24] & [25]) are made up of

- #1 variation of part characteristic due to production conditions (usually the *rate*)
- #2 variation of part characteristics over time in the field.

The demand noises (called *outer* noises in [24] & [25]) are made up of

- #3 customer duty cycles,
- #4 external environmental conditions induced by climate conditions and road inputs,
- #5 internal environmental conditions caused by complexity-induced interactions of neighbouring components, e.g. through transmitted heat and vibration.

Together, this collection is sometimes referred to as "the five noises". Robustness failure modes are caused by sensitivity of function to these noise factors, and occur when the demand placed on the design exceeds the capacity for some units, which is schematically illustrated in Figure 3.

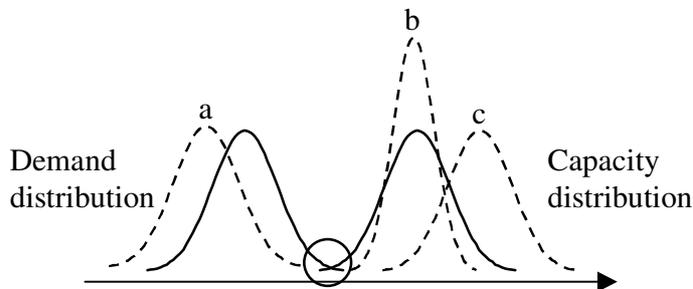


Figure 3: illustration of the interference between demand and capacity noises. The overlap is highlighted in the circle. Improvement to robustness can come about through a combination of three strategies; a) reduce the magnitude of demand, b) reduce the dispersion in the capacity around the average, or nominal value, and c) increase the average, or nominal capacity of the design.

Figure 3 motivates the idea of measuring the separation between demand and capacity. If the overlap between demand and capacity can be reduced, then this separation will be increased, and so position the design away from the point at which it might fail. However, we do not propose measuring this overlap directly (because it requires calculations related to the frequencies of the demand and capacity distributions, which are impossible to obtain, leading us into a similar dilemma regarding equation (1) in Section 7), but rather we exploit the knowledge of the physics, geometry, and material properties to propose a universal metric for the propensity of a design to fail, called the *distance from the failure mode*, as a way to measure the avoidance of failure modes due to robustness problems. Note also from Figure 3 that three strategies for robustness improvement emerge – reducing the demand (which can be achieved, for example, by configuring hardware to reduce noise #5), reducing the spread in capacity, and increasing the nominal value of the capacity.

In populations as large as automobiles, with the associated complicated noise space, failures caused by lack of robustness can never entirely be eliminated, but they can be minimized. We need to manage failures due to robustness problems, by understanding how the distance from the failure modes can be expressed as functions of the basic physics of the component, together with analysis of associated geometry and relevant material properties. These distances are measured relative to the noises, and can be adjusted by using the design variables at our disposal to increase the design capacity.

The idea of measuring robustness as a distance from a failure mode, rather than a probability, finally unifies reliability with the other attributes. We are now in a position to understand more with less.

10. The distance from the failure mode

We now discuss ways to measure the distance from the failure mode. Robustness, thought of as an interference between demand and capacity (Figure 3), can be formulated as a distance

measured with SI units, and captured by an appropriate metric which satisfies the following criteria; The metric

- can always be presented graphically
- always responds to noises
- measures the distance from the failure modes.

It is best to illustrate this idea directly with examples.

Firstly, we consider the Weibull plot of failure events. This is a convenient metric with which to start because it is also used extensively with the probabilistic definition of reliability, and so it forms a bridge to the definition of failure mode avoidance. Under the probabilistic definition the emphasis is on measuring aspects of the failure frequency, the *ordinate* of the plot. Under the new definition though, the emphasis is on the *abscissa* of the Weibull plot; this measures the location of the failure events relative to the life variable. If these failure events are induced by the appropriate noises, and if the useful life period can be equated to a point on the abscissa, a measure such as the "B₁₀ life" or some other quantile (e.g. see [16]) becomes a measure of distance from the failure mode. This idea is illustrated in Figure 4.

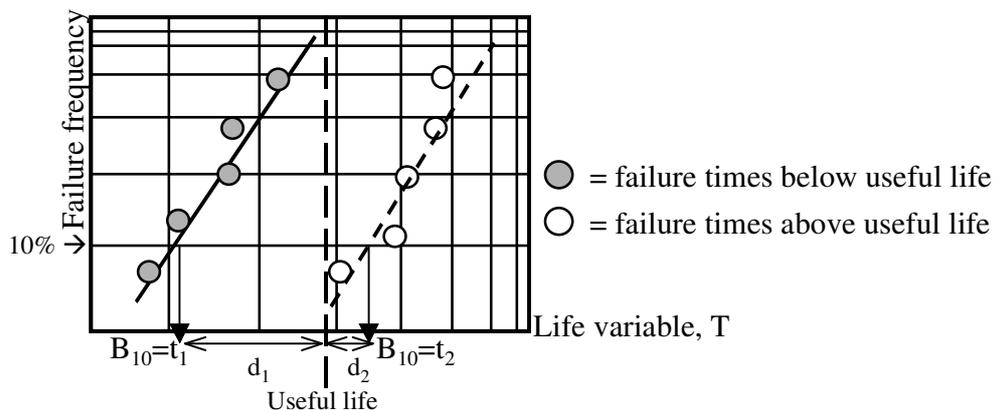


Figure 4: The Weibull plot as a robustness demonstration metric. The B₁₀ life (or some other quantile), measured relative to the useful life period, becomes a measure of the distance from the failure mode (denoted d_1 & d_2) provided the test that generated the data included the correct noises. The open circles illustrates a design with a greater distance from the failure mode (as evidenced by $t_2 > t_1$) and hence is more robust.

From Figure 4, we can see that the basic idea is to push the failure events as far away from the useful life as possible.

We now further develop the idea of a distance from the failure mode. Consider the case of measuring the variability induced in some physical characteristic as a result of exposing the system to noises. This physical characteristic has some nominal value, but will suffer variation due to the noises. Specimens that have a value lower than the nominal may suffer one kind of failure mode (e.g. there may be insufficient material for the part to work), and specimens that

have a higher value may suffer another kind of failure mode (e.g. there may be too much material). This variation could be measured with a standard deviation, a range, or some other measure of dispersion that possesses the *same units* as the data under investigation. Larger values of this dispersion metric will place some specimens closer to the failure modes, either above or below the nominal value. This idea is illustrated in the left-most picture of Figure 5.

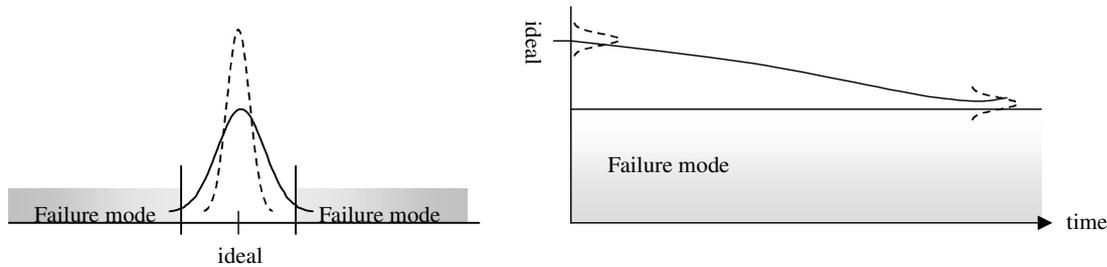


Figure 5: using a dispersion metric to measure the distance from a failure mode. The left most picture shows the distribution of component performance. That represented by the dotted line has a smaller measure of spread (e.g. as evidenced by a standard deviation), and is thus further from the failure modes. The right most picture illustrates the same idea, but with added degradation over time.

Often, characteristics such as those illustrated in the left most picture of Figure 5 may degrade due to continual exposure to repetitive demand. That is, the distribution may move toward the failure mode due to prolonged use. Measuring the extent of this degradation, by exposing the system to appropriate noises, is another way to measure the distance from the failure mode, as illustrated on the right of Figure 5. In these cases the extent of degradation deemed to be a failure might take some effort to derive, but the intention of course is to minimize the amount of degradation.

A further general schematic is illustrated in Figure 6, with the additional feature that the degradation is evaluated over a range of the input variable (in the sense of Figure 1).

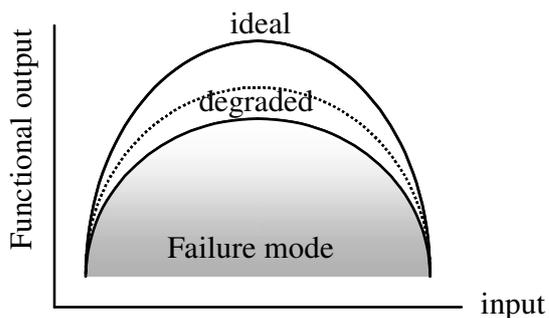


Figure 6: Degradation of a functional curve as a measure of the distance from the failure mode. The dotted line represents less degradation, and so is further from the failure mode.

As a final example, we show how the specific robustness metric advocated by Taguchi in [24] & [25], the "Signal-to-Noise" ratio, can be seen as a special case of a robustness metric that measures the distance from the failure mode. Taguchi's metric firstly requires that the relationship between the input and the output (see Figure 1) is linear, and passes through the origin. The motivation for this is based on measuring the amount of work done – if we put no

signal into the system, then the work done should be zero, and if the amount of input is doubled, the output should double. There are other requirements for Taguchi's metric to be valid, which are discussed in [26]. As with the previous metrics, this situation will be disturbed by noises, and the line will deviate from its ideal state. Excessive deviation leads to failure modes and so, again, this metric measures the distance from these failure modes as a function of the noises. This idea is illustrated in Figure 7.

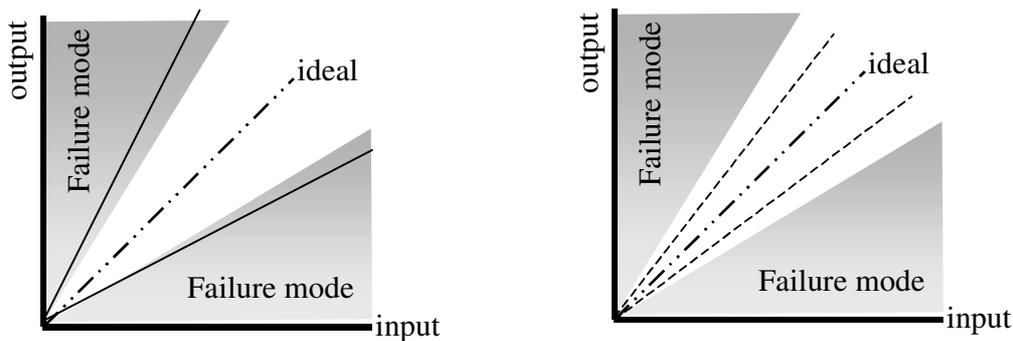


Figure 7: Graphical representation of Taguchi's signal-to-noise ratio to measure the distance from a failure mode, which are assumed to be two-sided. The picture on the right shows a system that is more robust (further from the failure modes) than the one on the left.

All of the metrics discussed above can be used to measure the effectiveness of a counter measure since a counter-measure will move the design further away from the failure mode (in each of the cases illustrated in Figures 4 through 7, this implies moving from the performances indicated by the solid lines to those indicated by the dotted lines). Note also, that the number of data-points needed to evaluate these metrics will be substantially less than the sample sizes typically required to estimate the probabilistic measure of reliability, where a sample size of 22 or more is often required (e.g. see [14], [15], & [16] for details on sample size calculations).

Finally, we remark that a related concept to those presented above is the idea of the Operating Window (OW) method, due to Clausing (see [19]). Essentially, the idea with the OW method, is that, rather than setting the engineering function as far away from the noises as possible, as we have seen illustrated in Figures 4 through 7, to move the region of the failure mode away from the current performance of the system. This idea is illustrated in Figure 8.

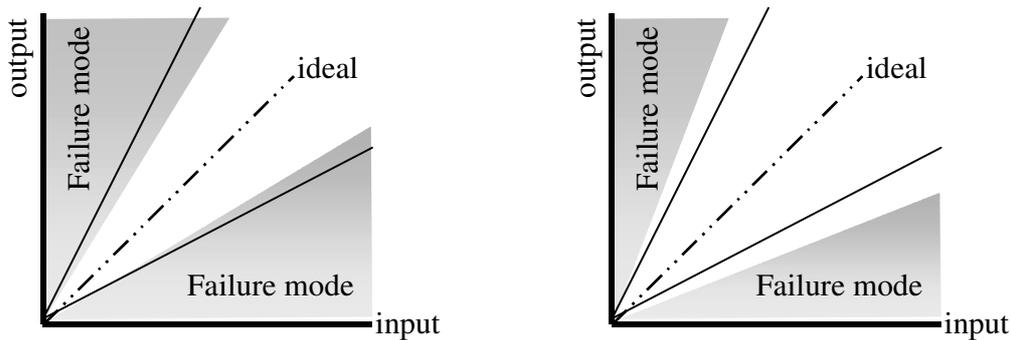


Figure 8: An illustration of the Operating Window (OW) method due to Clausing [19]. Note that the main idea is to move the failure modes as far away from the noise-induced performance of the system as possible, as seen on the right. This figure is a modification of Figure 7, but the OW method can also be demonstrated with any of the metrics illustrated in Figures 4, 5, & 6.

11. Transfer Functions

The counter-measures needed to move the design further away from the failure mode may be easy to deduce on occasion, but in many cases they will need to be discovered through a careful study of the way the component works, and an understanding of how design parameters affect the performance relative to the noises (including the interactions with neighbouring components in the system). The universal way to characterize this behaviour is through a transfer function, the mathematical representation of the laws of nature as they apply to the case at hand.

Determining the engineering actions required in formulating a counter measure to increase the distance from the failure mode calls for an analysis of the effect of noises on the function of the component. How can we use the idea of a transfer function to discover failure modes by analysis? For this, we connect Mischke's idea illustrated in Figure 1 to the concept of the "p-diagram", based on an idea due to Phadke [27, page 30], which followed the fundamental work of Taguchi [24, 25]. A general p-diagram is illustrated in Figure 9.

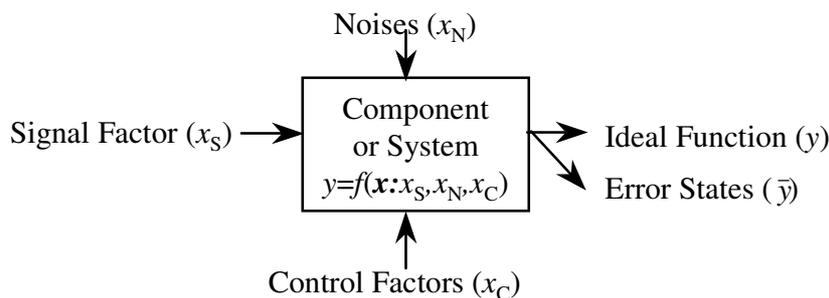


Figure 9: The p-diagram to describe an engineering system or component, as an extension to Mischke's idea (Figure 1). Note the addition of error states as compared to Phadke's original version in [27], which reflects noise type 5 (interactions of neighbouring components). The outputs (either y , the ideal state), or \bar{y} (the error states) can be expressed as a function of the inputs, $\mathbf{x}=(x_S, x_C, x_N)$

We now relate the concepts in Figure 9 to the basic diagram of Figure 1 and note that for any component there are three kinds of inputs (x 's) and two kinds of outputs (y 's).

The outputs are:-

- The *ideal function* – what we want the component to do, the ideal state of the output (denoted y);
- *Error states* - outputs related to unwanted conditions. These error states can lead to failure modes (which we will denote as \bar{y} , the complement of y);

while the inputs are

- The *signal factor* – the primary input to the system, the signal that makes it work (which we will denote x_S);
- *Control factors* – elements of the hardware, governed by physics, geometry and properties of materials, that can be manipulated in the design of the basic system (which we will denote x_C);
- *Noise factors* – disturbing influences that potentially disrupt the functioning of the system (which we will denote x_N , and were introduced in Section 9).

The question now is how to analyse for the presence of failure modes caused by lack of robustness, given the noises? Taguchi [24, 25] chooses always to detect the presence of failure modes (i.e. to measure the distance from the failure modes) through a Signal-to-Noise ratio (S-N ratio) as illustrated in Figure 7. In the context of the p-diagram, the S-N ratio has a useful interpretation, as follows; the signal factor (x_S) generates the ideal output (y), so that in an ideal case (taking x_C as fixed for now), x_S translates directly to y through a known relationship, while the noise factors, x_N , cause the presence of the error states, \bar{y} . With the correct interpretation, these error states can be thought of as variation in y (caused by the variations in x_N), so that $\bar{y} = \sigma_y^2$ say. The S-N ratio is then interpreted as the ratio of ideal output to unwanted output (y/\bar{y}), the effect of the signal divided by the effect of the noise²⁰. Making this ratio large will avoid the error states (and the consequent failure modes) caused by the error states, and so measures the distance from the failure mode.

In summary, transfer functions allow the discovery of countermeasures for the failure modes through analysis. These counter measures are developed by fabricating elements of x_C and managing elements of x_N to reduce the magnitude of \bar{y} so that the correct value for y is achieved for the correct input x_S .

11.1 Hypothetical example

To illustrate the idea of discovering counter measures to failure modes from transfer functions, we will look at two hypothetical examples. Suppose first that the transfer function is very simple, of the form

²⁰ Signal-to-noise ratio's are sometimes used in statistical analysis, for example the t-ratio.

$$y = \alpha \exp(\beta x_C), \quad (3)$$

where α and β are fixed parameters. We will suppose that the system does not depend on x_S , and further that x_C , the control factor, has a standard deviation, σ_C say, caused by production conditions in trying to achieve a nominal value for x_C of μ_C , say. Therefore the noise factor originates in the capacity space of the design. However, the distance from the failure mode is measured, not by σ_C , but by the standard deviation of y , σ_y (see Figure 5).

We can now use the seminal work of Jim Morrison, from 1957 [28] to connect σ_C with σ_y as follows:-

$$\sigma_y^2 = (\partial y / \partial x_C)^2 \sigma_C^2 \quad (4)$$

where the derivative is evaluated at μ_C .

In this hypothetical example, $\partial y / \partial x_C = \alpha \beta \exp(\beta x_C)$. If we take σ_C^2 as fixed²¹, we should clearly set a relatively small value for the nominal of x_C , since this then reduces the size of the derivative, and hence the size of σ_y . This idea is illustrated graphically in Figure 10.

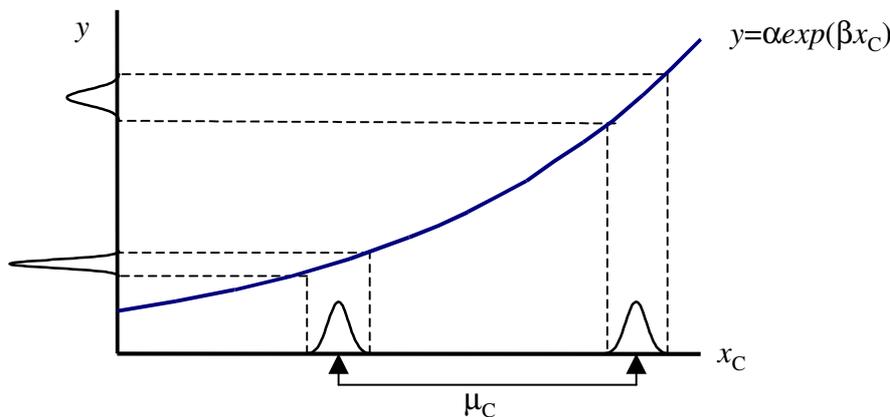


Figure 10: Illustration of the idea of transmitted variation, due to Morrison [29], exploiting the gradient of the transfer function. Note that as the nominal value (μ_C) of x_C is reduced, so is the magnitude of σ_y , in spite of the fact that σ_C remains constant. Since σ_y measures the distance from the failure mode, the counter measure is to reduce the value of μ_C .

Once the variance in y is minimized in this way, all that remains is to relocate the output to its correct target. Hence the counter-measure to increase the distance from the failure mode is to locate μ_C where the gradient of $y = f(x_C)$ is smallest.

As Morrison shows in [29], the result (4) easily extends to several x_C 's as follows:-

²¹ A simple extension allows σ_C itself be a function of μ_C , without much complication. A useful model is $\sigma_C = k \mu_C^p$. When $p=0$, σ_C is constant. When $p=1$, $\sigma_C \propto \mu_C$, a situation commonly found in practice.

$$\sigma_y^2 = \sum_i [\partial y / \partial f(x_{Ci})]^2 \sigma_{Ci}^2 \quad (5)$$

where the σ_{Ci}^2 are the associated production variances corresponding to each x_{Ci} .

The idea of using gradients of transfer functions to formulate counter-measures for failure modes can be extended to noise factors originating in the demand space. We will denote such a noise factor as x_N , and extend our hypothetical transfer function to include this noise as follows:

$$y = \alpha \exp(\beta x_C) + \gamma x_N + \delta x_C x_N. \quad (6)$$

Differentiating (6) with respect to the demand noise x_N yields a contribution to σ_y^2 of

$$(\partial y / \partial x_N)^2 \sigma_N^2 = (\gamma + \delta x_C)^2 \sigma_N^2,$$

where σ_N^2 is the variance of the demand noise. The value of σ_N^2 may be difficult to determine, but to minimize its impact on σ_y^2 , we should set $x_C = -\gamma/\delta$ so that the gradient in the x_N direction is zero²². This extends Morrisons original idea to include demand noises. A treatment for robustness against demand (or outer) noises was probably first discussed by Michaels in 1964 [29]. Of course, in some situations, the counter-measure dictated by the demand noise may be in conflict to that being suggested to counter the capacity noise. In such a case, the trade-off should obviously result in the minimum possible value of σ_y . If this resulting standard deviation still places the design too close to the failure mode, other counter-measures to nullify the effect of the noise factors, such as removal of the noise, or compensation devices, will need to be tried.

Figure 11 is a graphical representation of the analysis just discussed for equation (6).

²² Another way to see this is to re-arrange the terms in (6) involving x_N as $(\gamma + \delta x_C)x_N$. Setting $x_C = -\gamma/\delta$ reduces the coefficient of x_N to zero.

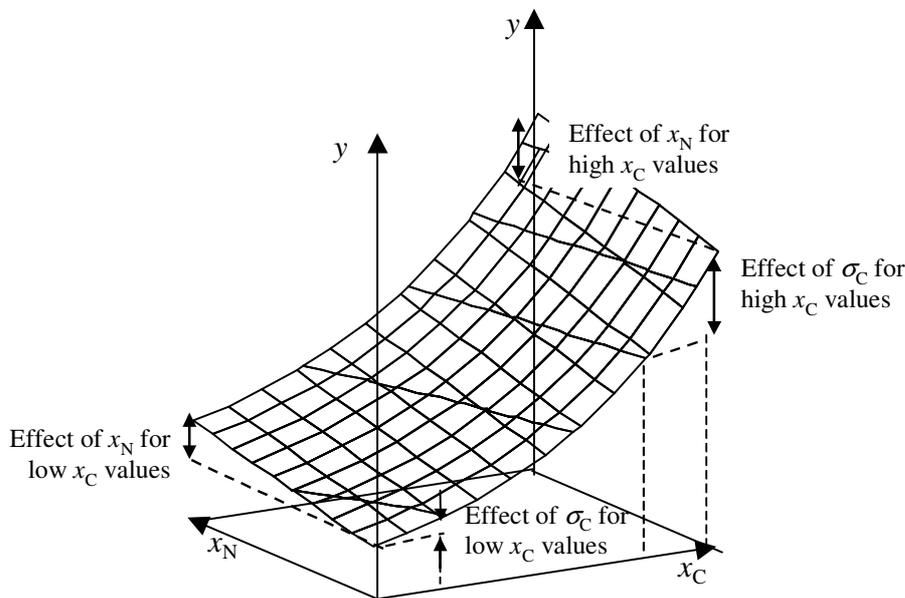


Figure 11: A response surface showing the shape of the transfer function (6) in x_C and x_N . Note that in this example, reducing the nominal value of x_C minimizes the impact of both the production variation around the nominal and concurrently the effect of the demand noise x_N , since the gradients in both directions are lower for smaller x_C values.

11.2 Real Examples

We now turn our attention to real examples of failure mode avoidance, all taken from applications within the Ford Motor Company. Although in some of these examples the failure modes were not discovered until the product was in the field, we present the analysis in such a way as to demonstrate how the failure modes *could* have been identified well before the design was finalized, and therefore how the symmetry of the engineering process (2) could have been preserved on the sense of Section 8.

Sticking spring in a push-push switch (hard failure)

The ability of a spring to perform its function in a push-push switch depends on the Hooke's constant of the spring being able to overcome the friction caused by interference with materials inside the mechanism. Material shrinking due to extreme cold temperature in the field increases this friction. Therefore, the effect of temperature can be expressed as a function of force, the force becomes the demand noise in the sense of Section 9.

Translating temperature, which is experienced at the vehicle level, to a force, which is what the component sees, is an application of systems engineering. We are recognising that the vehicle environment can be translated to an equivalent environment at the component level. Recall that we have already made the observation that it is only components that fail. This idea can be nicely illustrated with a so-called "V" diagram in Figure 12. This picture also implicitly illustrates the idea of symmetry developed in Section 8.

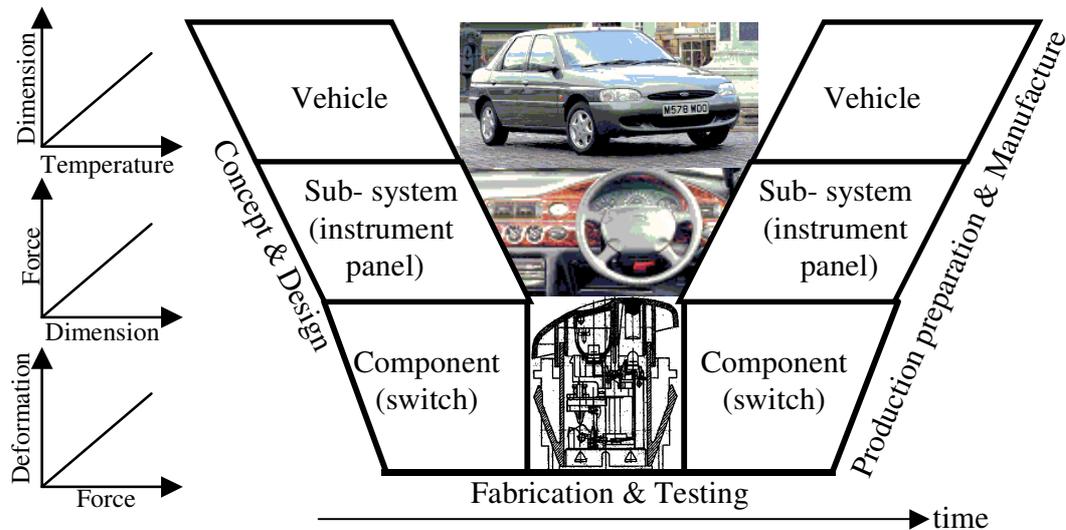


Figure 12: the "V" diagram, with reference to the switch example. Note that the failure mode that will be observed in the vehicle can, with some thought, be exposed at the component level, by cascading the effect of the noise factor (temperature) to the component (illustrated in the transfer functions on the left of the diagram). The idea of symmetry is captured in this picture – the design should look the same on the right hand side of the "V" as it does on the left.

It is vital to understand how the components work in the larger system, so that counter-measures can be taken early - that is the countermeasure for the failure mode is taken on the left hand side of the "V", not the right hand side. This will then preserve symmetry.

We use the material properties of the instrument panel to translate the external temperature into a change in dimension, which in turn is translated into a force, causing the distortion of the switch housing, and so increasing the friction coefficient inside the mechanism, causing the switch to stick. The propensity to fail can be measured as a distance and therefore predicted as a function of the force (and hence the temperature, which is the cause of the failure mode in the field). Figure 13 illustrates two designs, A and B, measured on this metric.

Design A is the initial design, released to the field, and found to fail in cold temperatures. This design has a negative distance from the failure mode compared to design B, which includes the countermeasure of making the switch housing stiffer to resist deformation and hence makes the design robust to temperature. This solution was favoured over increasing the stiffness of the spring (which would have raised the dotted line in Figure 13), because another failure mode related to "switch feel" would have been introduced.

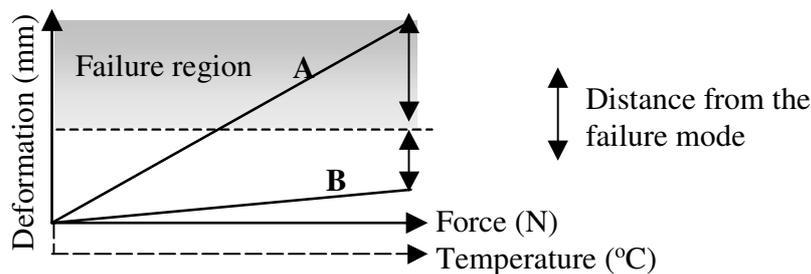


Figure 13: Two switch designs A & B. There are two counter measures for the failure mode of a sticking switch evident in this picture. Either the dotted line could be raised, thus avoiding the failure mode in design A; this would be equivalent to increasing the Hooke's constant of the spring in the switch. Or the gradient of the relationship between force (and hence temperature) and deformation could be reduced, as illustrated by design B; this is equivalent to making the housing around the switch stiffer, thus resisting deformation due to temperature.

Customers in the field discovered this failure mode initially, but the analysis of Figure 13 shows that it could (indeed should) have been discovered while the switch design was still in the development phase. Once the failure mode has been discovered, and the physics and geometry understood, it is clear what counter-measures need to be taken. Taking these counter measures during the development phase means they do not disrupt fabrication, production planning, and manufacture, thereby preserving symmetry.

Excessive stowage times for seat belts (soft failure)

In the United States, the JD Power study [13] indicates that the time for a seat belt to retract into the mechanism for stowage in the "B-pillar" when exiting the vehicle is the second highest failure mode in terms of frequency (about 7% of customers complain at 3 years in service).

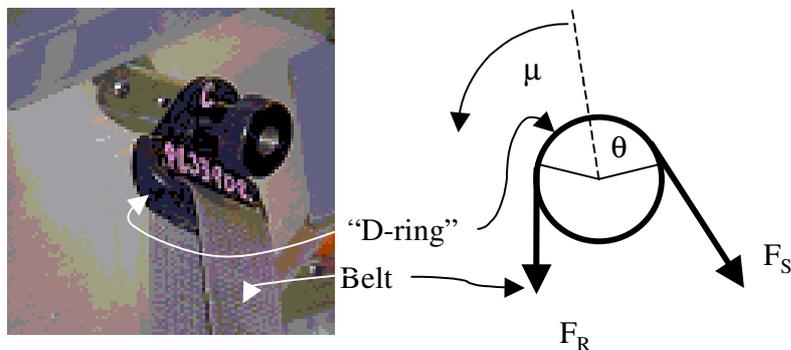


Figure 13: schematic of the forces involved in stowing a seat belt over the cross-section of the "D-ring" (illustrated on the left).

A development of the transfer function with stowage time as the output is needed to understand what is causing this failure mode, so that counter measures can be taken. Figure 14 illustrates the geometry involved – the seat belt is shown extending over a cross section of the D-ring. The basic force equation given, for example, in the Bosch Automotive Handbook [30, page 49], shows that $F_S = F_R \exp(-\mu\theta)$. Since $F_S = mg$, where m is the mass of the seat belt mechanism, and $d = \frac{1}{2}gT^2$, where d is the length of the belt being stowed, we derive the following equation relating the stowage time, T , with the coefficient of friction, μ ;

$$T = \sqrt{(2md/F_R)} \exp(1/2\mu\theta) \quad (7)$$

Note that this transfer function is a special case of our hypothetical example (3). Equation (7) clearly illustrates that counter-measures to excessive stowage times involve reducing the magnitude of μ , or increasing the magnitude of F_R (usually by a stronger retraction spring in the mechanism). The friction coefficient μ is determined by the properties of the materials of the D-ring and the seat belt webbing. Since stronger springs have other failure modes (e.g. pressure on the occupants chest during use of the belt), we explore counter-measures based on adjusting μ .

During manufacturing, μ will vary from part to part because of production conditions associated with the rate. This in itself will result in variation transmitted to the retraction time, through the transfer function (7), according to Morrison's theory (4). An added complication in this example is that the value of μ drifts in the field due to usage, a fact ascertained through inspecting and measuring seat belts in vehicles with various mileages. This increase in μ is caused by contaminants building up on the surface of the webbing material and on the D-ring. The effect on the retraction time is illustrated in Figure 15.

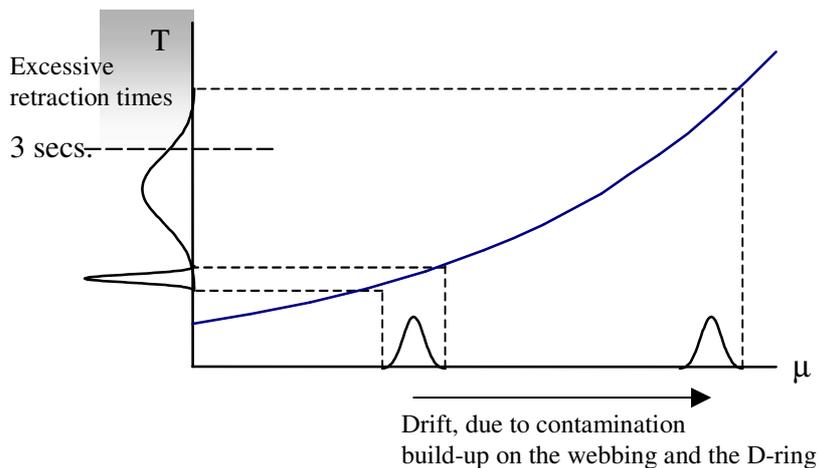


Figure 15: the effect of drift in the coefficient of friction, μ , on seat belt stowage times, T . Studies in the field with customers indicate that stowage times in excess of 3 seconds cause significant dissatisfaction.

Figure 15 shows that counter measures to excessive stowage times of seatbelts can be achieved by firstly setting a target for μ in the lower range, where the gradient of the transfer function is reduced. This will ensure minimal variation in stowage times from vehicle to vehicle. And secondly, choosing materials for the D-ring and webbing that resist the build up of contaminants during use that will cause μ to increase and then reside under a very steep part of the transfer function. The steep gradient will result in some customers, whose coefficient drifts to the extreme, suffering excessive stowage times of their seat belts²³.

²³ The problem of excessive seat belt stowage times affects many vehicle manufacturers in the United States. A subsequent study of vehicles in the field showed that the vehicle manufacturer with the lowest reported failure rate for excessive stowage times did indeed use materials for the D-ring and the seat belt webbing that resisted contaminant build up.

The illustration of the failure mode in Figure 15 shows a classic robustness problem. How can such failure modes be prevented in the future? Design guidelines can now be created which set a low target for μ , and specify standards for materials that resist contaminant build-up. In future, it would be expected that designs for seat belts must meet this standard; designs that do not can be regarded as mistakes, since we now know what to do to avoid this failure mode.

Variable torque in an electric motor (soft failure)

This case study is reported briefly in [12]. The torque (T) generated by a certain type of electric motor used to raise and lower glass in power windows, is a function of the wire diameter (D) of the windings, the length of the wire in one wrap (L_D), the wire resistivity (R_W), the length of the armature (L_A), the length of the magnet (L_M), the internal diameter of the motor housing (f_D), magnet thickness (M), rotor core diameter (p_D), magnet angle (θ), and a magnet material constant (ρ). This set of parameters makes up x_C . Each of these parameters has a potential to be adjusted (within feasible ranges) to decrease the sensitivity of the design in the presence of production variation, according to the result (5). Failure modes occur if the time taken to raise and lower a window is too long or too short, and so we can measure the distance from these failure modes by σ_T , the standard deviation of the torque caused by production variations around the nominal values of the parameters in x_C .

Table 3 shows each of these parameters with their nominal values, and standard deviations expected at production rates of about 1 piece per minute.

| x_C | Diameter of wire - mm (D) | Length of wire on one wrap - mm (L_D) | Resistivity - $\Omega m^2/m$ (R_W) | Length of armature - mm (L_A) | Length of magnet - mm (L_M) | Interior diameter of housing - mm (f_D) | Magnet thickness - mm (M) | Rotor core diameter - mm (p_D) | Magnet angle - deg. (θ) | Magnet material constant (ρ) |
|--------------------------------------|-------------------------------|---|--|-----------------------------------|---------------------------------|---|-------------------------------|------------------------------------|----------------------------------|-------------------------------------|
| Nominal range | 0.50 0.55 | 100 120 | 0.80 0.85 | 32 38 | 45 50 | 38 42 | 4 5 | 25 30 | 130 145 | 400 450 |
| standard deviation (σ_{Ci}) | 0.0033 | 1.166 7 | 0.0007 | 0.366 7 | 0.333 3 | 0.016 7 | 0.0667 | 0.016 7 | 0.666 7 | 2.5 |

Table 3: ranges for nominal values and standard deviations for parameters, x_C , in the electric motor.

The Transfer function for torque in terms of the variables in Table 3 is given by

$$T = kD^2L_AL_M\theta\rho M[L_DR_W(f_D-p_D)]^{-1} \quad (8)$$

where $k=2.5 \times 10^{-4}$ is a constant. We will suppose, for the illustration here, that the target value for T is 4Nm. There are of course an infinite number of nominal values of the variables in x_C that give the answer 4Nm, but the question is, which one yields the smallest value for the standard deviation in torque, σ_T ? Hence our problem is this; how do we minimize the standard deviation σ_T given by (5), subject to the equation (8) hitting the target of 4Nm? This problem can be

solved with any optimization software (we used here the Solver function in Microsoft® Excel®), and the following nominal values were obtained (Table 4).

| x_C | Diameter of wire - mm (D) | Length of wire on one wrap - mm (L_D) | Resistivity - $\Omega\text{m}^2/\text{m}$ (R_W) | Length of armature - mm (L_A) | Length of magnet - mm (L_M) | Interior diameter of housing - mm (f_D) | Magnet thickness - mm (M) | Rotor core diameter - mm (p_D) | Magnet angle - deg. (θ) | Magnet material constant (ρ) |
|---|----------------------------------|--|--|--------------------------------------|------------------------------------|--|----------------------------------|---------------------------------------|-------------------------------------|--|
| Nominal | 0.55 | 120 | 0.80 | 38 | 50 | 38.7 | 5 | 26.5 | 145 | 450 |
| $\partial T/\partial x_{Ci}$ | 14.545 | -0.033 | -5.000 | 0.105 | 0.0801 | -0.328 | 0.800 | 0.328 | 0.028 | 0.009 |
| $(\partial T/\partial x_{Ci})^2(\sigma_{Ci})^2$ | 0.0023 | 0.0015 | 0.0000 | 0.0015 | 0.0007 | 0.0000 | 0.0028 | 0.0000 | 0.0003 | 0.0005 |

Table 4: nominal values for parameters for the electric motor. These nominal values result in a torque of 4Nm from (8), and a standard deviation of $\sigma_T=0.0988$, obtained by summing the numbers in the last row according to equation (5), and taking the square root.

If we are not satisfied with the resulting standard deviation for torque (in the sense that some units will be too close to the failure modes), the analysis of Table 4 also tells us which parameters are the best candidates for planning counter-measures through tightened tolerances – these being, in order of priority, M , D , L_D and L_A , since these factors contribute most to the value of σ_T .

Excessive starting times for engines (soft failure leading to hard failure)

Our final example concerns avoiding failure modes of too long start times for engines (taken together with the seat belt and power window examples, it is clear that customers of automobiles do not like to waste time waiting for features of their vehicles to function). We give just a brief treatment here since a detailed account of this study is given in [26].

Customers are concerned when the start time of their engine exceeds about 1 second – significant reductions in satisfaction and confidence occur for starting times longer than one second. The main noise factor that contributes to this failure mode is cold weather.

Since, as we have seen, failure modes need to be found early in product development, waiting until engines are assembled in prototypes to evaluate starting times is too late. What needs to happen is that the fundamental principles of achieving a good start time are characterized in the component hardware. To this end, the primary requirement to achieve a good start time is to ensure that the mixture, y , of fuel to air is just right in the cylinder when the spark plug ignites the mixture. This ratio can be set at a nominal condition by injecting a given quantity of fuel – this is the signal factor, x_S , in the sense of the p-diagram of Figure 9.

The transfer function for how this ratio behaves is quite easy to determine – since the amount of air is constant, the more fuel that is injected results in this ratio increasing by a corresponding amount – this is just geometry. However, the noise factor of temperature, x_N , causes some disruption to this state. When the temperature is cold, the air density increases so that the fuel to air ratio decreases (the mixture becomes lean); when the temperature is warm, the air density is reduced and the fuel to air ratio increases (the mixture becomes rich).

This phenomenon can be represented by the following transfer function;

$$y = \alpha_0(1 + \alpha_1 x_N) x_S \quad (9)$$

where α_0 and α_1 are parameters to be determined. Mixtures that are too lean or too rich result in engines that require a longer time to start. The distance from these failure modes is measured primarily by the parameter α_1 in (9). Figure 16 illustrates the geometry involved, and shows the metric used to measure the distance from the failure mode.

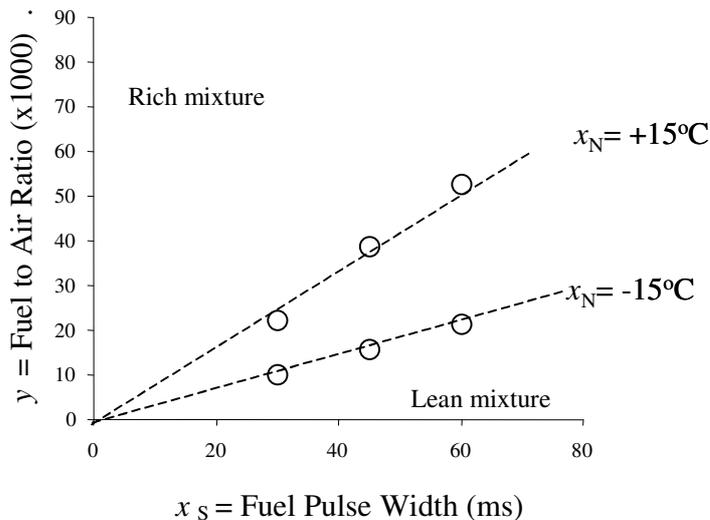


Figure 16: Fuel to air ratio's in an engine, as affected by temperature. Mixtures that are too lean or too rich cause ignition problems leading to excessive start times.

The engineering challenge now is to determine engine hardware that can be used as a counter measure to too much dispersion between the two temperature lines in Figure 15. Small values of α_1 will indicate a design that provides this counter measure, and so we require a formulation $\alpha_1 = f(x_C)$ that will help us locate this design. Unlike the situation we faced in the previous examples though, it is difficult to deduce from first principles or known theory, exactly what this transfer function is, and hence what such a configuration of hardware for the counter-measure might be. We therefore resort to discovering this transfer function through experimental induction using the powerful technique of statistical experimental design. The book by Grove and Davis [31] gives a thorough overview of the theory of experimental design, with many automotive examples. Essentially, what is required in this example is to experiment with variables that make up the space of x_C , to find a combination that reduces the dispersion in Figure 16, as measured through α_1 in (9). The set of control factors, x_C , in this example consists of variables related to characteristics of the fuel injectors, together with geometrical adjustments (in time and space) of the location of the injectors and spark plug.

The complete experimental set-up, transfer function development, and analysis are reported in [26] and so will not be repeated here. The results of such an approximating transfer function allows a prediction to be made as to where, in the space of x_C , the most effective counter measure

to the failure mode resides. These predictions are illustrated in Figure 17, which shows the metric for the initial development engine configuration (prior to the transfer function being developed), and the resultant metric once the optimal configuration had been derived through analysis of the estimated transfer function.

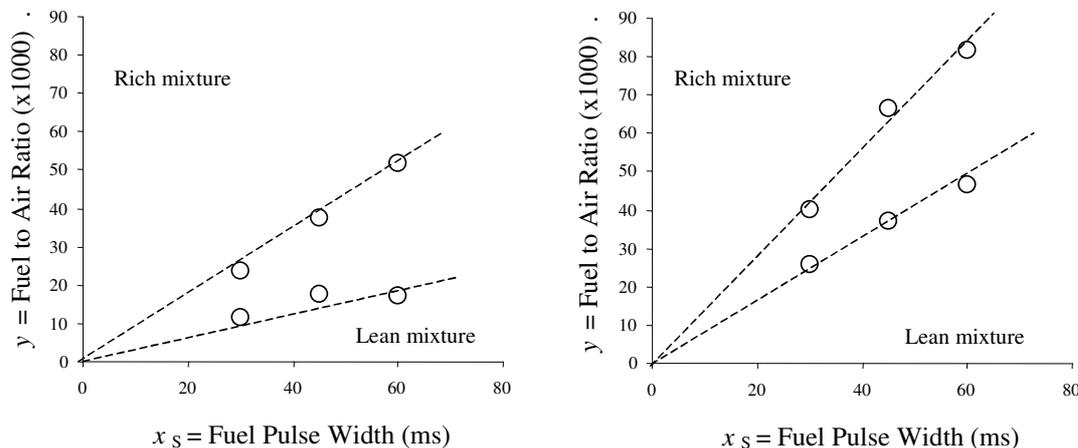


Figure 17: distances from the failure modes of lean and rich mixtures. The plot on the left is the engine configuration prior to analysis of the transfer function, and the plot on the right. Is after the resulting counter-measure has been developed by the fitting of an empirical transfer function through experimental design

As part of this study it was discovered that a heater device, which was fitted to some injectors as a counter-measure for the cold ambient, was not required. Thus, failure modes related to over-design of the hardware (which might well have resulted in more engineering changes later in the development phase of the project) where also avoided, and the hardware remained as parsimonious as possible.

The avoidance of the failure mode of too long start times were verified by building this engine configuration into a prototype vehicle and testing it in various temperature conditions – this test illustrated the low variation of start times across the temperature range, as required. Subsequent field data with vehicles in the hands of real customers also confirmed that failure modes associated with excessive start times in cold ambient conditions had been avoided, as detailed in [26].

12. Summary

The scientific principles of symmetry, parsimony, and unification provide an important framework in guiding an approach to engineering based on finding and eliminating failure modes early in the development of products. Interplay between induction and deduction is an important element in implementing these scientific ideas within engineering, as the case studies have illustrated.

In automotive engineering, the inductive-deductive iteration is immediate, but complicated by the presence of product, production, and customer-induced variability (collectively described by

the noise factors). Statistical science provides a framework for dealing with this variability; illustrated in this paper are two important statistical ideas – the concept of transmitted variation, and the use of experimental design to obtain approximating transfer functions for the underlying, unknown state of nature. We call the use of statistics with engineering in this way, *statistical engineering*.

Reliability, one of the most important attributes to customers, requires an approach based on physics, geometry and properties of materials, rather than the traditional approach using probability. This is motivated by the fundamental idea of unification, and consequently the definition of reliability as failure mode avoidance is appealing from the standpoint of parsimony. This change of emphasis re-directs effort away from measuring the frequency of *when* things fail, to understanding *how* and *why* they fail. The failure mode, rather than the frequency, then becomes the fundamental quantity to be managed, which leads to the idea of preserving a form of symmetry in engineering (that is, the final design in production looks the same as the one on the engineering drawing).

Finding failure modes early in product development is key to achieving reliability in the field, because early discovery of failure modes allows for early development and adoption of counter-measures. Since a failure mode only has to be found and fixed once, this avoids making changes to the design late in development, which ultimately results in failures occurring in the field. Finding failure modes through testing can be most effective when this testing is focussed on exposing components to the right noises. Finding failure modes through analysis is greatly enhanced by the use of transfer functions, the mathematical representations of how the laws of nature apply in each case. Additionally, with reliability measured as a distance, in SI Units, from the failure mode, rather than a failure rate or probability, sample sizes in testing can be greatly reduced.

Based upon the principles of symmetry, parsimony, and unification which govern the character of physical law, it would seem compelling that the definition of reliability as failure mode avoidance should replace, forthwith, the definition based on probability.

13. Acknowledgments

This work started as a seminar presentation at the Ninth ASA/IMS Spring Research Conference on Statistics in Industry and Technology at the University of Michigan, Ann Arbor, MI, USA, May 21, 2002. Sir David Cox, FRS, was in the audience and he encouraged me to write up the material. This paper is the result. Ed Henshall, Vasily Krivtsov, and Chris Gearhart, all of the Ford Motor Company, made important comments on an earlier draft, which led to improved presentation. Don Clausing, formerly professor of mechanical engineering at MIT, provided some very profound insights, which lead to greater clarity in the first half of the paper. Deborah Ashby, professor of medical statistics at Queen Mary College, London, helped with the discussion of the analogy between engineering reliability and medical survival.

14. References

- [1] Mischke, CR. *Mathematical model building – an introduction to engineering*. ISU Press, 1980. ISBN: 0 8138 1005 1
- [2] Gauch, HG. Jr. *Scientific method in practice*. CUP, 2003. ISBN:0 521 01708 4
- [3] Mahon, B. *The man who changed everything – the life of James Clerk Maxwell*. John Wiley, 2003. ISBN 0 47086088 X
- [4] Ulrich, KT, & Eppinger, SD. *Product design and development*. (3rd edition) McGraw Hill, 2004. ISBN: 0 07 247146 8
- [5] Kane, Thomas R. & Levison, David A. *Dynamics: theory & applications*, McGraw-Hill, 1985. ISBN: 0 07 037846 0.
- [6] Box, GEP. "Science and statistics". *Journal of the American Statistical Association*, Vol. 71, No 356, 1976, pages 791-799.
- [7] Cox, DR. in "Deming and the pro-active statistician", by GJ Hahn, *The American Statistician*, Vol. 56, #4, 2002, pages 290-298.
- [8] Jaynes, ET. *Probability theory – the logic of science*. CUP, 2003. ISBN 0 521 59271 2.
- [9] Carver, RP. "The case against statistical significance testing", *Harvard Educational Review*, Vol. 48, #3, 1978, pages 378-399.
- [10] Joyce, HR. "Beyond reasonable doubt". +*Plus Magazine*, Issue 21, Sept 2002. (see <http://plus.maths.org/issue21/features/clark/>)
- [11] Farmelo, G. (Ed.). *It must be beautiful – great equations of modern science*. Granta Books, 2002. ISBN 1 86207 479 8.
- [12] Parry-Jones, R. *Engineering for corporate success in the new millennium*. Royal Academy of Engineering, Westminster, London. 1999. ISBN: 1 8716 3483 0
- [13] JD Power and Associates. *Vehicle Dependability Study*, 2003, USA.
- [14] Greene, AE. & Bourne, AJ. *Reliability Technology*, John Wiley, 1972. ISBN: 0 471 32480 9.
- [15] Lewis, EE. *Introduction to reliability engineering*. John Wiley, 2002. ISBN: 0 47180989 6.
- [16] Meeker, WQ, & Escobar, LA. *Statistical methods for reliability data*. John Wiley, 1998. ISBN: 0 471 14328 6
- [17] Feynman, RP. "Personal observations on the reliability of the shuttle". Appendix F, *Report of the presidential commission on the space shuttle Challenger accident*. William P. Rogers (Chair), US Government, Washington DC, 1986.
- [18] Davis, TP. "The fallacy of reliability prediction in automotive engineering". *Automotive Excellence*, 1998, pages 19-21.
- [19] Clausing, D. "Operating window – an engineering measure for robustness" *Technometrics*, Vol. 46, #1, 2004, pages 25-29.
- [20] Krivtsov, VV, DE Tananko, and TP Davis. "A regression approach to tire reliability analysis", *Reliability Engineering & System Safety*, Volume 78, number 3, 2002, pp 267-273.
- [21] Davis, TP. "Reliability improvement in automotive engineering", in *Global Vehicle Reliability – prediction and optimization techniques*; JE Strutt & PL Hall, Eds., 2003, Professional Engineering Publishing. ISBN: 1 86058 368 7
- [22] Clausing, D. *Total quality development – a step-by-step guide to world class concurrent engineering*. ASME Press, New York, 1994. ISBN: 0 7918 0035 0.

- [23] Davis, TP. "Symmetry in engineering". *In preparation*.
- [24] Taguchi, G. *System of experimental design*. UNIPUB/Kraus, 1987. ISBN 0 527 91621 8.
- [25] Taguchi, G. *Introduction to quality engineering – designing quality into products and processes*. Asian Productivity Association, 1986. ISBN: 92 833 1084 5.
- [26] Davis, TP. "Measuring robustness as a parameter in a transfer function", in *Reliability and robust design in automotive engineering*, Society of Automotive Engineers, Inc., Warrendale, Pennsylvania, USA. 2004-01-1130. ISBN: 0 7680 1380 1.
- [27] Phadke, MS. *Quality engineering using robust design*. Prentice Hall, NJ, 1989. ISBN: 0 13 745167 9.
- [28] Morrison, SJ. "The study of variability in engineering design". *Applied Statistics*, Vol. 6, #2, pages 133-138, 1957.
- [29] Michaels, SE. "The usefulness of experimental design (with discussion)". *Journal of Applied Statistics*, Vol. 13, #3, pages 221-235, 1964.
- [30] Robert Bosch, GmbH. *Automotive Handbook*, 4th edition. Robert Bosch, GmbH, 1996. ISBN: 1 56091 918 3.
- [31] Grove, DM, & Davis, TP. *Engineering, quality, and experimental design*. Longman, UK. 1992. ISBN: 0 582 06687 5.