



Delft University of Technology

Engineering Laboratory Experiments—a Typology

Zwart, Sjoerd

DOI

[10.1080/19378629.2022.2126783](https://doi.org/10.1080/19378629.2022.2126783)

Publication date

2022

Document Version

Final published version

Published in

Engineering Studies

Citation (APA)

Zwart, S. (2022). Engineering Laboratory Experiments—a Typology. *Engineering Studies*, 14(2), 158-182.
<https://doi.org/10.1080/19378629.2022.2126783>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Engineering Laboratory Experiments – a Typology

Sjoerd Zwart

To cite this article: Sjoerd Zwart (2022) Engineering Laboratory Experiments – a Typology, Engineering Studies, 14:2, 158-182, DOI: [10.1080/19378629.2022.2126783](https://doi.org/10.1080/19378629.2022.2126783)

To link to this article: <https://doi.org/10.1080/19378629.2022.2126783>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 17 Oct 2022.



[Submit your article to this journal](#)



Article views: 62



[View related articles](#)



[View Crossmark data](#)

Engineering Laboratory Experiments – a Typology

Sjoerd Zwart

Faculty of Technology, Policy and Management, Technical University, Delft, The Netherlands

ABSTRACT

With the introduction of large commercial industrial laboratories at the end of the nineteenth century, many types of experiments were institutionalized that do not aim at testing hypotheses. This paper builds a *typology of experiments* in techno-science, by analysing more than two hundred and fifty real-life technical projects. This resulted in *four testing types* (tests of hypotheses, of designs, of means-end knowledge, and of models or software), *three determining types* (developing working principles, preferred actions, and determining values of variables or relationships between variables) and one trial-and-error *type of pure exploration*. The typology is developed by working back and forth between thick descriptions of the experiments including their goals, and the development of *six criteria of differentiation*, to wit: determining versus testing; measurement scales of (in)dependent variables; intrinsic versus instrumental value of the outcomes; proximate function of the outcome; distant role of the outcome in the embedded project; the descriptive or normative character of the proximate or distant outcomes. The typology opens up inspiring methodological and philosophical research questions.

ARTICLE HISTORY

Received 28 July 2021

Accepted 15 September 2022

KEYWORDS

Laboratory experiment; thick descriptions of experiments; levels of description; typology; ideal types of experiments; theory-driven; exploratory; epistemic and action-guiding experiments; criteria of differentiation; engineering; hierarchical project plans

Introduction

The appearance of commercial laboratories was an integral part of the second industrial or technical revolution. In 1876, Thomas Alva Edison, founded one of the first laboratories for research and commercialization of electrical technologies in Menlo Park, New Jersey; he employed more than hundred researchers.¹ The importance of Edison's initiative lay not only in the production of amazing new products, but also in the creation of an institution whose goal was the development and commercialization of inventions.² With the establishment of research in commercial laboratories came also the institutionalization of various types of engineering experiments. This article sets itself the goal of categorizing these types.

Many philosophical works on experimentation by, for example, Roger Bacon, John Stuart Mill, Auguste Comte, Claude Bernard, or Pierre Duhem, already recognize other functions

CONTACT Sjoerd Zwart  s.d.zwart@tudelft.nl  Department VTI, Faculty TPM, Delft, University of Technology, Postbus 5015, Delft 2600GA, The Netherlands

 Supplemental data for this article can be accessed here. <https://doi.org/10.1080/19378629.2022.2126783>

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

of experiments than simply testing hypotheses. Summarizing this work, Franklin and Perovic discuss some of them in physics.³ They mention: choosing between two competing theories, calling for a new theory, confirming or refuting a theory, providing evidence for the existence of an elementary particle involved in an accepted theory, or that determines the mathematical form of a theory. They take sides with Ian Hacking claiming that experiments have ‘a life of their own’, independent of theory.⁴

Note, however, that all these functions ultimately concern the creation of *descriptive knowledge*. This seems natural, at least at first sight, because Franklin and Perovic concentrate on physics. Despite the colossal technological influence on experiments in physics today (e.g. CERN’s Large Hadron Collider), they do not consider the much richer variety of experiments in technology.

Developments in thermodynamics, electromagnetism, and fluid dynamics during the Second Industrial Revolution made scientific and engineering practices increasingly interdependent.⁵ Experimental scientists achieved spectacular technological results and engineers contributed to scientific breakthroughs. Thus, to structure the science-technology relationship, Zwart and de Vries recommend to concentrate on *types of research projects* rather than theories or practices.⁶ This helps to distinguish the different roles of experiments, as working packages in an engineering project have a role-related hierarchical means-end structure.⁷ These roles, in combination with their characteristics, establish the type of an experiment. The goal this article sets itself is to provide a typology of experiments encountered in the practices of engineering projects, which serves epistemic and methodological purposes.

The structure of the paper is as follows. In the next section, we first discuss what is meant by ‘experiment’; next, we elaborate the different levels on which experiments can be described; then our method and its justification are explained, and at the end the criteria of typology are discussed. The third section is dedicated to the main empirical content: the eight types of experiments are introduced by means of examples. In the fourth section, the types are compared with some distinctions already found in the literature: theory-driven versus exploratory experiments (Friedrich Steinle) and epistemic versus action-guiding ones (Sven Ove Hansson). And the last section briefly indicates possible methodological and philosophical follow-up questions.

What is in this paper the difference between a ‘classification’ (or ‘taxonomy’) and a ‘typology’? In general, a classification or taxonomy is a partition of a set, where a partition is a collection of mutually exclusive and collectively exhaustive equivalence classes. That is, every entity belongs to one and only one class. Consequently, the boundaries of the classes are unequivocal. The present typology violates this mathematical condition. It classifies experiments less strictly using the notion of *Weberian ideal types*. Weber defined ideal types as methodological instruments, which were non-normative mental constructions, one-sidedly accentuating the essence of a concrete aspect of a phenomenon.⁸ Constructing an empirical typology comes down to finding distinctive ideal types and their characteristics that cover many similar entities in the field. Regarding laboratory experiments, for instance, Boyle’s experiments make the ideal of the hypothesis-testing type. All the eight types in our typology are characterized by a model example. Discussions can arise about the type of a specific experiment or about whether a new type should be introduced. This objection does not, however, invalidate our typology any more than it would the typology of colors, which evokes the same discussions.

The success criteria of a typology consider its comprehensiveness (does it cover almost all entities); the conceptual clarity (do most entities clearly belong to one type, without evoking many discussions); and its balance (a small number of types should not cover almost all entities). Note that, in contrast to descriptive beliefs, typologies are not (approximately) true or false; they advise how to cut up reality for some purpose that might be closely related to truth and falsity (biological classification, periodic table) but need not (typology of sciences or of speech acts).

Preliminaries

Before creating a typology one should identify the objects to be classified. Here, the question arises whether an experiment includes its results. If not, the experiment is only an action protocol. This is not problematic when the results of the experiments are (almost) the same. What, however, if with the same protocol, the outcomes differ substantially? Then, either the experiments are the same but underdetermined, or they are different. Additionally, if an experiment is identified with its protocol, it can only fail due to a deviation from its protocol, such as a technical failure. To evade this and similar complications, here, an experiment is identified with its protocol and its direct result. The second identification issue, the various possible levels of description, will be dealt with after the next section.

Characterization of 'laboratory experiment'

In this article, a laboratory experiment is a token execution of a means-end designed subproject embedded in a larger techno-science project with the purpose of learning something. It should therefore be reproducible, which means that the execution of the same protocol produces similar effects. The learning should take place by systematically manipulating the most important input factor(s) of the assumed underlying mechanism, and controlling all other relevant aspects as much as possible. Finally, the effect(s) on the output (the results) should be observed or measured. When almost all tokens of the same experimental protocol culminate in comparable results, they form an *institutionalized experiment type* such as, e.g. the Wheatstone bridge measurement of an electrical resistance.

Let us consider some consequences of this characterization. First, neither pure computer simulations nor thought experiments are considered real experiments, unlike their experimental validations. Next, to enable learning, repeated experiments should give similar outcomes and should not be 'unique'. Because experiments must be partially controllable, large unique 'social-technical' trials are not experiments as characterized above. Also phenomenological observations and measurements, although necessary ingredients, differ from experiments because of the control and manipulation requirements. Since the design of an action plan determines group intentions, experiments can be performed by individuals or by groups. Finally, the result of an experiment can concern the truth-value of some statement, or the suitability of a design or an intervention for some goal.

Due to the characterization above, intentions of researchers play a crucial role in the identification of experiments. Regarding artifacts, we usually distinguish between their physics, the material or molecular structures, and the purpose for which the artifact was conceived and developed. Only the molecular arrangements fail to characterize an artifact because of the lack of intentionality. The molecular arrangements of its stones, wood

and glass does not make a construction a cathedral, which requires the intentions of its makers and users. In this article, the same is assumed for laboratory experiments. Consider Heinrich Hertz' experiment with the Ruhmkorff coil designed to show the existence of Maxwell's EM-waves. Hertz is claimed to have said about these experiments: 'It's of no use whatsoever. This is just an experiment that proves Maestro Maxwell was right – we just have these mysterious electromagnetic waves that we cannot see with the naked eye. But they are there'.⁹ Some years later, also using the Ruhmkorff coil, Guglielmo Marconi carried out similar experiments to develop wireless telegraphy and radio emissions. Again somewhat later, Ruhmkorff coils were used in experiments to ignite internal combustion engines. Despite the physical similarity of these three types of experiments, they have to be classified differently because of their different purposes.

Let us conclude this section with some semantical remarks. First 'means-end knowledge' is contrasted with 'working principle'. The first is considered general action-knowledge of the form: if one wants to achieve technical goal *G* within engineering context *C*, one is advised to carry out engineering action *A*. The second is connected to a unique design or model. It is a causal chain specifically designed to make a technical artifact meet its requirements. Next, a 'proof of concept' is experimental evidence to show that some chosen working principle performs as anticipated. Finally, in a 'feasibility study' researchers determine whether a technical solution is economically feasible.

Levels of description

Let us turn to the different levels-of-experiments descriptions. The first and thinnest description level only considers the experiments' phenomenology. On this level the inputs, outputs and actions are only described in the way cameras capture images without considering any intentions or goals whatsoever. These phenomenological descriptions typically use direct observations: pictures, lights moving, needles being close to some number on a dial, stretching springs, human movements, etc.

On the second level, descriptions grow thicker when formulated in terms of changing variable values and explanations of what is physically happening with the objects during the experiment. Statements like 'the force is five Newton' are already interpretations because we only have direct access to displacements. Even at this physical level, however, an experiment fails only if it does not follow the protocol. Michelson, in his attempt to show differences between the speed of light rays due to their difference in velocity relative to the aether, could write to Lord Rayleigh: 'The expected deviation of the interference fringes from the zero should have been 0.40 of a fringe – the maximum displacement was 0.02 and the average much less than 0.01'.¹⁰ But without his intentions he could not have claimed that his experiment had failed. It is this physical level of description, which remains close to the physical skin of the mechanism that may lead to the belief that most laboratory experiments are of the same type. Indeed, most laboratory experiments can be described at this physical level, in terms of changing values of (in)dependent variables and controlling intervening variables. Those descriptions are not false, but they neglect important differentiating methodological experimental details.

On a third level of description an experiment is conceptualized as an action, and the physical description is extended with the proximate goal explicated in the embedding project plan. On this level (and the next) hypothesis testing and design performance

can be distinguished when the physics of both experiments are the same.¹¹ Consequently, a typology of experiments starts to be feasible on this level. In addition, only here, experiments may fail even if the protocol has been followed meticulously. Thus, because the aim was to measure the speed of the aether wind on earth, the Michelson-Morley experiment may be considered to have failed although it was carried out in accordance with its protocol. Failed experiments in this sense are those where nature (artificial or designed) refuses to answer, or gives answers completely outside the scope of the expectations.

Michael Bratman explains how all human plans are hierarchical.¹² This also holds for Techno-science projects. Consequently, on the final description level, the interpreted outcome of the experiment is also considered as a means for its embedding project goal. Laboratory experiments then become sub-plans in an overarching research design plan. They are considered as work packages in a larger research project, which, besides the proximate goals, also provides for more distant purposes of the experiment. The distant goal of an experiment is the role that this experiment plays in its embedding project. Within our Michelson-Morley example the well-established corrective conclusion 'Aether wind does not exist' belongs to the description on project plan level. From this comprehensive point of view the experiment was an important starting point for further developments.¹³ 'Thin experiment' descriptions refer to those on the first two levels, and 'thick descriptions' to those adding proximate goals and distant purposes. Our ideal-type typology of experiments is based on thick experiment descriptions.

The method and its justification

The purpose of this article is to offer an approach to cutting up the reality of engineering experiments for epistemic and methodological purposes. This is conceptual research, which requires different methods and validation than purely descriptive empirical projects. Let us consider the paper's methodology and its validation.

One way of designing a typology of experiments is to turn to the literature. The advantage of consulting well-documented examples is that they provide clear conceptual anchoring points. The method has two drawbacks, however. First, most of the existing experiment descriptions are written from the perspective of descriptive knowledge production, and therefore run the risk of being biased. Second, only following the literature one would be in danger of missing out experiment types that serve the purposes of engineers. To counter these two threats this article follows a bottom-up approach and carries out original empirical research.

The empirical material built upon is a database of short reports of more than a thousand Bachelor End Projects (BeP) carried out at the mechanics faculty of the Delft University of Technology (3mE) between 2002 and 2015. In these six-months projects, four to six students, closely supervised by high level staff, worked on engineering assignments to obtain their Bachelor's degrees. Researchers of all departments were asked to hand in open-ended assignments with unknown answers. BeP results were regularly published, even in refereed journals. The large majority of BeP reports contain experiment descriptions. The author has been involved in teaching methodology parallel to the BePs and in hands-on coaching of at least five hundred groups. The typology below has been developed by investigating

all experiments performed in more than two hundred BePs carried out between 2012 and 2015. As the 3mE departments use many experimental strategies also used elsewhere in the TU Delft it is assumed that the database gives a sufficiently rich and perhaps even comprehensive overview of most experimental actions of engineers in (academic) practice.

The development of an ideal typology comes down to *working back and forth* between the empirical data (the experiments) on the one hand, and the type descriptions (characteristics and criteria) on the other. At the start were some generally accepted characteristics (e.g. hypothesis testing, and exploration) or cases (Boyle, Edison, Froude) which determined the first version. Going back to the database, problems were encountered such as, for instance, with its comprehensiveness or lack of discriminatory force (here e.g. differences in scales of the independent and dependent variables). Path-dependence necessitated a new start as *different levels of description* and *scale differences* were not taken into account. After some back-and-forth movement the typology and the criteria started to stabilize and the other experiments received their natural placement. At the end the typology was comprehensive and practical, at least as far as academic engineering experiments were concerned.

The three most apparent threats to the (construct) validity of the proposed typology are the following. First, what if the samples taken from the 3mE department (TU Delft) are biased? It is unlikely that staff would advise students to perform other types of experiments than they do themselves. Perhaps other faculties at TU Delft provide different types or would have completely different distributions of types? That would pose a threat to the content validity. Indeed the generalization of the present typology might be an interesting follow-up research question, especially for other practical sciences such as the medical and the agricultural ones. It should be noted that this typology is concerned with the types of experimental *actions* and not experimental *contents*, since for the latter the threat of bias is larger. If one extends the scope even further, one might fear a bias toward academia. Perhaps in a commercial setting, engineers would perform other 'quick and dirty' experiments for lack of time or money. However, it seems unlikely that the latter would lead to other types of experiments, rather than sloppier executions.

The second threat to validity is the clarity of the 'type' definitions and the usability of the criteria. The chosen way to counter this threat to criterion validity is to publish it in the exploratory phase and see what colleagues think. If the typology is extended with statistical recommendations, it can also be released as a survey to practitioners to assess its practical value.

Subjectivity is the third apparent construct validity threat for the typology. Would attempts of other researchers to construct a typology with the same methodological and epistemological purposes converge with the one presented here? This is a complicated issue that requires at least another paper to be answered satisfactorily. First, one must ask whether convergent validity is an appropriate requirement, since it seems to vary from one subject to another. The periodic table encompasses more empirical input and converges more than psychological typologies, or even biological taxonomies. In addition to true (or false) claims about experiments, typologies also involve the choice of criteria and their relative values, which can be established in a variety of ways. Valuations are not true or false, but must prove practical, provide more or less methodological guidance, or provide epistemological insight. Second, the writings on experimentation do show at least some convergence. We saw above that completeness requires that functions other than just

hypothesis testing be considered. Moreover, the present typology bears some resemblance to the proposals of Steinle, Burian, and Hansson (see below).

Criteria

What criteria did our back-and-forth work bring up? First, the most general criterion distinguished between experiments that *determine* and experiments that *test*. While exploring, constructing or discovering, the former develop new claims or determine preferred actions, artifacts or material properties. Examples include the determination of coefficients of expansion, or stress–strain ratios of alloys, or the drag–lift ratio of a wing or propeller. Experiments of the second kind validate hypotheses, artifacts, or procedural knowledge, such as Pasteur’s experiments on spontaneous generation, or the *Greyhound* experiment by which Froude tested his method of predicting the friction of a ship’s hull. Almost all experiments are either determining or testing.¹⁴

The second criterion is the *type of scales* of the dependent and independent factors or variables; especially the difference between nominal and ordinal scales is significant. This distinction emerges already on the second thin level of description. The third criterion is whether the experimental *outcome also has intrinsic value* besides its instrumental value for the embedding research project. Some experiments only have instrumental value; others have additional intrinsic value even outside the context of the research project in which it is carried out. The fourth criterion is closely related; it considers the *proximate function* or role of the experimental outcome. Fifth, also the distant purpose of the experimental outcome in the *project plan* may play a role. Finally, the sixth criterion is whether these proximate or distant outcomes have a *descriptive or normative character*. Most of these criteria were identified during the process of categorization, which was carried out against the background of general considerations such as: taking experiments as action plans and distinguishing goals of engineering projects. The resulting types are not simply blind combinations of the six criteria mentioned, but are coherent ideal types to be found in the practices of laboratory life.

As an illustration, consider an experiment in the database designed to determine whether it makes sense for pathologists to humidify human bone while sawing it, because of differences in spread of wet and dry particulates. On the level of thin descriptions, the expected outcome can be taken as a test of a descriptive hypothesis with nominal independent and ordinal (or even quantitative) dependent variables. In our typology, it will be taken to be testing the consequences of a practical manipulation because of its minimal intrinsic but important instrumental value for pathologists.

Proposed typology of laboratory experiments

This section presents the empirical heart of the present paper. It introduces the eight types of experiments by presenting examples. The boundaries between the types are sometimes fuzzy, and for some experiments the chosen type may provoke discussion. This does not invalidate the typology. The fact that people may disagree about the name of one specific shade of red does not deprive our color system of its value. The order of presentation is didactical and not systematic. Every type is characterized by means of the criteria in Table 1, and associated with a famous historical experiment, which functions as its heading and paradigmatic example.

Table 1. The criteria of experiment differentiation.

1	Determining versus Testing
2	Measurement scale types of (in)dependent variables
3	Intrinsic versus instrumental value of the outcomes
4	Function or role of outcome in the embedding research/project plan
5	Type of distant outcomes in this plan
6	Descriptive or normative proximate or distant outcomes

Boyle type – hypothesis testing

Let us start with best-known type of scientific experiment, which occurs regularly in technological contexts as well. The experiments of this type test a hypothesis, which identifies a quantitative mathematical, or ordinal relationship between the independent and dependent variables. Note that the outcome or conclusion of these experiments is descriptive and as value-free as possible.

The example from the database concerns the behavior of a hanging liquid column within a research project that investigated the drinking mechanisms of cats. For the experiment, such a column was produced by a circular disk raised from a liquid surface with a uniformly accelerated speed of approximately 4.5 m/s^2 . It was driven by a linear actuator. The disc continued to move upward as long as the liquid column, formed by the disk, did not break. The dependent variables in this experiment were the maximum volume V_{max} of the liquid column, the time at which the column broke t_k , and t_{max} the time at which the volume of the column was at its maximum. A high-speed video camera (500 images/s) recorded the column. The hypothesis read: 'An increase of viscosity leads to an increase in V_{max} , t_{max} and t_k '. It turned out that for V_{max} and t_k the hypothesis holds whereas it fails for t_{max} .

This project concerned the descriptive knowledge regarding a drinking mechanism and the experiment had at least as much intrinsic as instrumental value. Moreover, it measured the (in)dependent variables on quantitative scales although the hypothesis compared them on ordinal level. This makes this experiment an example of the (ordinal) hypothesis testing type.

Characteristics and digression

Let us call experiments testing a hypothetical quantitative relationship between variables Boyle-type experiments. They have the following characteristics. First, the intention of the experimenters is to test a mathematical relation between (a few) measurable independent and dependent variables. Consequently, the independent variable(s) should not be only nominal – i.e. unordered cases, which is a characteristic of the practical experiments to be discussed below. Second, although the outcomes of the experiments may have instrumental value for other goals in the embedding research project, typically, Boyle-type experiments have intrinsic value. Experiments of this type test whether a proposed mathematical relationship between variables correctly describes the natural or artificial world in the sense of Robert Boyle's famous words: '[T]he Hypothesis, that supposes the pressures and expansions to be in reciprocal proportion'.¹⁵

Let us take an interesting philosophical digression. The outcomes of most pure Boyle-type experiments are indifferent to causal directions; they remain neutral regarding the causal influence between the independent and dependent variables. Often methodological handbooks require that the independent (variable) states have to cause the

dependent (variable) states. Addition of mercury to one leg of a one-sided plugged U-tube will cause an increase of pressure of the parcel of air under the plug. In contrast, the dilution of air by means of a piston-cylinder device causes a reduction in the air pressure of the remaining air. But taking two experiments with opposite causal directions as confirmation of one pressure-expansion hypothesis makes the causal direction of that hypothesis redundant. And yet, it is this position that Boyle takes in the marvelous fifth chapter in his 1662 book. The mathematical relation between pressure and volume is considered the same under and above one atmosphere and the reverse underlying causal directions in the two experiments are completely irrelevant. Boyle concatenates the two experiments into one causally 'directionless' experiment, whose outcome is the hypothesis cited above. Galilei's mathematical relationship between pendulum frequency and length is also of the directionless Boyle-type. A ratchet that would impose one specific frequency would 'cause' the length of a pendulum that keeps swinging in phase.

Directionless experiments are a main ingredient of the scientific revolution. As Hansson rightly claims engineers were carrying out quantitative experiments long before the seventeenth century.¹⁶ The unique developments during the scientific revolution, however, were not only the application of experiments to scientific questions, but mainly the replacement of cause–effect theory by direction-free mathematical manipulation. The distant experimental outcomes were mathematical relationships between variables, apt for pure mathematical treatment, the results of which could again be tested in other directionless experiments. This was the revolutionary development and prepared the way for 'pure' physics, which, fed by experiments, was detached from cause–effect directions and had, as far as we know, no precursor in any preceding civilization. Its products do not feature causal (or other types of) directions, with perhaps the exception of thermodynamics. Directionless experiments have not been carried out by engineers before the scientific revolution.

Wright flyer type – design testing

In contrast to cases where researchers want to test hypothetical relationships between (in)dependent variables, in design projects experiments are carried out to find out whether the designed artifact functions according to expectations. One example in the database of this type was in the project of miniaturization of liquid pumping devices. A glass piezoelectric pump was built with a membrane of polydimethylsiloxane (PDMS). This membrane was activated with polyvinylidene fluoride (PVDF), a strong piezoelectric, which was glued on the membrane. Due to alternating electric tension the membrane would alternate between a convex and concave form, thus changing the volume of the 10 mm diameter pump chamber. On this small scale the construction of valves is difficult. Therefore, the inlets and outlets of the pump chamber were made cone-like; the outlet had the cone's small diameter side, and the inlet its large diameter side connected to the pump. This nozzle-diffuser design is known to have valve-like functionality. An experiment was carried out to validate the prototype. It turned out that the single layer of PVDF changed its form sufficiently only under high electric tension (240 V). Moreover, the addition of another PVDF layer did not make the membrane deflect sufficiently. It was concluded that reducing the glue layer or finding a stronger piezoelectric material would solve the problems.

Characteristics

Experiments testing designs similar to that just described will be put under the heading of *Wright Flyer test experiments*. Although Orville and Wilbur Wright did not test their designed artefacts in a laboratory, the images of them lying on their motorized gliders perfectly portray the test-design character of this type of experiments. They are all set up to test whether a unique design behaves as expected and fulfills the design requirements. If not, they are used to find out how to improve the design.¹⁷ Their function in a hierarchical research plan is to evaluate a unique blueprint, prototype or design. Typically, the intrinsic value of the direct measurements, if any, is small, and contrasts with their large instrumental value. Based on these experiments it is decided how the design process should be continued. Consequently, the distant conclusions are strongly normative. Even if, for instance, the project concerns the design of a measurement instrument, which is tested by comparing all measurements results in the entire range against measurements standards, then still this is not a Boyle-type of experiment, because of the lack of intrinsic value of that comparison.

Note that scientists also use testing design experiments when they test a measuring setup or apparatus. Nevertheless, considered by their outcomes and goals, testing designs remains engineering. These goals are never the description of a mathematical relation between variables, which we encountered in the previous section. Thus, physicists use test-design experiments for their instrumental and not intrinsic value.

Swan/Raven type – determining the preferred practical intervention

Besides hypothesis and design testing, also a third group of experiments has been identified in the recent literature. These experiments concern discovering the consequences of some type of intervention or manipulation and to put the outcomes in an order of preference. These practical experiments come close to what Hansson calls ‘direct action-guiding’ experiments.¹⁸

The real-life example from the database appeared in a project regarding steerable instruments used during minimally invasive surgery in which instruments were developed with steerable tips (50 mm), actuated by extra cables to enhance maneuverability. The question was how to raise the bending and the torsional (rotational) stiffness of the tip. To that end, tips were constructed with parallel cables, and two with helix cables, one with a half and another with an entire revolution along the 50 mm axis. The tips existed of three concentric layers of cables. The inner layers of all three consisted of parallel cables, which enabled steering of the tip, and the outer layers rotated around the longitudinal axis for bending and torsional stiffness. The cable layers were constructed on a flexible core, and the stiffness cables were clamped in cable fixations at the beginning and end of the tip. Measurement of these tips showed that the bending stiffness of the half-helix tip was higher than that of the tip with the parallel cables, and that of the tip with one-helix cables. Moreover, the torsional stiffness increased with the number of revolutions. It was concluded that half-helix cables were most suited for steerable tips. They featured the largest bending and torsional stiffness and avoided unpredictable displacement behavior displayed by other one-helix tips. Note that not the proximate measurements but the more distant objective determines the type of this experiment.

Characteristics

Practical experiments in engineering (and other intervention sciences) are typically about determining the consequences of technical intervention or manipulation within well-defined circumstances. Their role in a hierarchical research plan is to find out the 'optimal', or most preferred intervention in that context; they discover means-end knowledge. Typically, the independent variables of practical experiments are nominal (e.g. different types of interventions, artifacts, or situations) and the dependent variables may be quantitative, but the precise values are often relatively unimportant; it is the order of the quantities to be 'optimized' that really count. Note that because no statistic measures the correlation between nominal and quantitative (or ordinal) variables even the most approximate outcomes of practical experiments are distinctively different from Boyle-type experiments. Although the direct quantitative outcomes have only comparative descriptive importance, the more distant outcomes typically have normative impact. Despite the subsidiary importance of their actual quantitative results, practical experiments usually have substantial intrinsic value.

In 1867–1868 William Froude conducted practical-like experiments when he famously compared the resistance of the *Swan* and *Raven* ship hulls in 1867–1868. He showed on three different scales that sharp ship hull models (*Raven*) produced more friction than the blunter ones (*Swan*) at higher speeds. The pure comparison of two models regarding their friction at various speeds can be taken as the image of practical experiments.¹⁹ All practical experiments are therefor brought together under the heading *Swan/Raven* type of experiments.

Note that a practical experiment characterized as determining the 'optimal' intervention or manipulation differs from choosing the appropriate unique working principle some design. The former is much more general and not geared toward one specific application as a unique design. In practice, however, the borderline between a practical experiment and finding a design working principle may sometimes be fuzzy.

Practical experiments are often conducted in situations where scientific theories fail to provide definitive answers and the results of such experiments are more or less satisficing rather than true or false. The choice of a ship hull depends on more than only its friction at high speeds.

Determining parameter values or relations between variables

The practical experiments of the previous section play the role of determining the intervention or manipulation with the most desirable consequences. The experiments in the present section determine something else, viz., the value of parameters or the relationship between variables. They do not explicitly test a hypothesized relationship, which makes the difference with Boyle-type of experiments. If the experiment serves a distant descriptive purpose, it typically precedes hypothesis testing. If, however, the outcomes of the experiment clearly serve a practical distant purpose, the experiment aims at the development of means-end knowledge. Boyle-type of experiments do not need to consider possible collateral damage, as engineering projects typically do when their goal is means-end knowledge.

The example was part of a project to develop a new body-powered arm prosthesis. It concerned an experiment designed to determine the pinch forces needed to hold rectangular and cylindrical objects with a hook made by TRS Prosthetics. In this experiment the

object size (27 and 44 mm for cylindrical and 14 mm, 27 and 44 mm for rectangular objects), weight (varying between 0.1 kg and 1.2 kg in steps of 0.25 kg) and bending stiffness were the independent variables; the required pinch force was the independent one. Every measurement was repeated ten times. The pinch force was measured indirectly via the force on the cable that powered the prosthesis. For the three object forms, five weights, and degrees of stiffness the necessary pinch forces were measured. It appeared that for the 14 mm object the pinch force was 1.7 ± 0.3 times lower than for 27 and 44 mm objects. Moreover, rigid and relatively stiff objects needed almost the same pinching. Only objects with considerably lower stiffness needed a smaller pinching force. As an extra, the researchers found out that the TRS hook needed more pinching force than predicted by the Coulombs law of friction.

Although the proximate outcomes of this experiment bear similarity to the mathematical relationship testing of Boyle experiments, their intrinsic value is much less than that of the latter. To know that the smallest pinching force to hold a rectangular object of size 44 mm with weight 0.875 kg is 30 N is uninteresting in itself. For the designers of a body-powered arm prosthesis, it is however very important. This newly developed knowledge of quantitative relationships clearly serves the distant purpose of design knowledge. The proximate outcomes are however descriptive and become normative in relation to the more distant purposes.

Characteristics

Experiments set up to find the value of parameters or the relationship between quantitative variables are concerned with determining rather than testing relationships. The independent and dependent variables are quantitative and the proximate purpose is descriptive. If the intrinsic value of the outcomes is high (e.g. determination of the speed of light, or the relationship between speed and rolling friction), then these experiments play a constructive part in the hierarchy of a descriptive knowledge plan. Note by the way that, because of the iterative character of the empirical cycle the borderline between these experiments and quantitative hypothesis testing may sometimes become fuzzy.

If the intrinsic value of the outcomes is clearly less important than the consequences for means-end knowledge (or even a design concept like skin friction, or a drag/lift parameter, etc.), then the experiment plays a normative rather than a descriptive role in the embedding project hierarchy. Consequently, experiments of this category may occur in project plans aiming at descriptive or means-end knowledge. Depending on their role in these plans the outcomes of the experiments may be more or less context dependent. The more the dependent variables are interpreted normatively, the more similar the experiment becomes to a practical means-end experiment.

The experiments in this section that have enough intrinsic value to become independent descriptive knowledge will be of the *stress/strain* type. These are experiments that aim at determining a value of a parameter or of the mathematical relationship between variables. If, however, such a quantitative experiment misses this intrinsic value and only becomes relevant for the achievement of another, more distant aim, then it will be called a *Froude's Torque* type of experiment. To determine the skin resistance due to viscosity induced by shear forces, and not wave making, Froude measured the friction of many vertical boards of different lengths (and sharp edges to evade wave making) moving under water. This resulted in Froude's formula of viscosity friction, which he used to predict the resistance of

large new ship hulls. It remained secret until far after his death – another clear difference with the swift publication of Boyle's law.

Systematic parameter variation

According to Walter Vincenti, *Systematic Parameter Variation* (PV) is a 'procedure of repeatedly determining the performance of some material, process, or device while systematically varying the parameters that define the object of interest or its conditions of operation'.²⁰ According to the PV method, first all relevant variables must be identified, then they are systematically varied and, finally, the results of the dependent variable(s) are presented in an orderly fashion in tables. The normative, more distant results (e.g. the highest lift/drag coefficient or the lowest resistance, etc.) are closely connected to these descriptive, proximate results. In 1901, the Wright brothers (bicycle repairers by trade) famously designed the first systematic wind tunnel PV experiments. Another well-known and more recent example of PV is the Wageningen B-screw series (1930s–1970s), in which the torque and thrust of 120 propeller models are measured (ranging between two and seven blades), at different speeds, and for blade-propeller area ratios between 0.3 and 1.05. The results are still being discussed in modern handbooks on marine propulsion.²¹

PV is often applied for technical purposes when a scientific theory is insufficiently developed or is unable to provide reliable and accurate quantitative answers to often urgent technical questions. Although most (or perhaps all) causal influences (relevant independent variables) on the dependent variable (the effects) are known, the situation might simply be too complicated for precise scientific elaboration. Typically, the independent variables mutually influence each other to such an extent that isolating them is impossible and assess their individual contribution to the change of the dependent variable. PV is a subcategory of the experiments covered in this section.

Interestingly, we will see below that Steinle regards PV as a characteristic of Exploratory Experimentation.²² Scientists, for whom by definition the distant goal is discovering aspects of some world structure, also apply PV to get a grip on new phenomena to be explored, which often leads to fundamental new conceptualizations. Hansson, however, considers PV to be possibly a 'generalization of the control group' and as such a 'safeguard in experimental design'.²³ In section 5, we will come back to this claim. Apparently, PV is a method rather than a type of experiment for which the distant goal determines its type.

Edison's filament type – determining a new and unique working principle

Experiments determining a unique working principle belong to a hierarchical plan for a specific design. They endeavor to find out whether this principle will work for the specific design. Let us look at an example where two working principles were compared.

Force control in robots is about how to let the robot apply the required amount of torque (force) with its end-effector to the outer world. In the search for cost-effective force control of a specific RoboCup robot, two promising concepts were compared. Concept A calculated the external force by comparing the motor torque, measured by the motor current, with the torque predicted by the dynamic model of the arm. Concept B measured the torques (forces) between the robot's hand and arm. These forces were decomposed into horizontal and vertical components, which enabled the robot to produce

the required forces with its end effector. To implement concept A, the kinematic model had the effector make straight up-and-down movements along a line, where the angles and velocities were fed into the dynamic model, which calculated the torque forces in the joints. These forces were also measured by the motor current. The differences were considerable and unpredictable. Thirty different values of the motor current (spread 5 Nm) related to 1 Nm of expected torque. This was not repairable. Concept B was first implemented with a fixed set-up, measuring two forces for eleven rotated positions (between pointing upwards toward pointing downwards) of the hand with known weight. With an affine-transformation method the two forces were decomposed in forces in the x - and y -direction. After the load cells had been calibrated, this method enabled the robot to determine the direction and magnitude of the external forces, and using the kinematic model, to move its hand in the force direction. A proof of concept had been successfully established.

Characteristics

The experiments just described were about determining working principles for some specific design. Their independent variables are, on the one hand, the experimental set-up to be implemented in the design, and on the other, the often quantitative values of its crucial variables. The most proximate outputs are typically quantitative (descriptive) values. However, these are so close to the normative design requirements that the outcomes of the experiment may be considered to be a proof of concept, or a failure (or a suggestion about how to improve the working principle). Because this describes the function or role of the experiment in the embedding project plan, most of the experimental value is instrumental. If compared with the experiments that help to decide about the most appropriate means for some technical end, developing means-end knowledge is typically more general than specific proofs of concept. The dividing line between determining a working principle for a unique design and testing a design may seem thin. However, it should not be overlooked; the former involves verifying whether a given principle will be able to provide the required functionality for a specific design, while the latter involves testing whether a design prototype will indeed produce the expected results.

A paradigmatic image for finding an appropriate working material is the frantic way in which Edison sought the appropriate filament for his electric light bulb between 1878–1880 at Menlo Park. This finally resulted in Edison's famous patent of November 1879 for a light bulb with a carbon filament. Generally, experiments that try to establish working principles often result in patents, if successful, something that tests of hypotheses almost never do. Interestingly, Edison's experiments with light bulbs also led him to observe the 'Edison effect', the emission of electrons from heated metals. He observed that an extra foil in the bulb connected to the negative potential compared with the filament, did not induce a current between foil and filament, whereas it did if connected to the positive potential. Later, John Ambrose Fleming used this effect as the working principle for the first vacuum tubes to amplify electrical signals, thus providing the impetus for modern electronics.

Froude's Greyhound type – testing means-end knowledge

Technical laboratories have been around for more than one century and a half, and many institutions of engineering education have been transformed into universities. Still the

emancipation of engineering means-end (m-e) knowledge experiences much opposition. That all m-e knowledge is reducible to descriptive knowledge,²⁴ or that m-e knowledge is nothing more than following a recipe are popular objections to taking academic m-e knowledge seriously. In contrast, many laboratories' files display experimental testing of m-e knowledge geared toward technical action and support of engineering m-e knowledge. Let us consider an example from our database.

In a materials-science project, a method to determine the plastic zone size around fatigue cracks was tested. In the proximity of such a crack, the local stress is higher than the yield stress that produces a plastically deformed zone. The deformations produce dislocations and bending in the crystal lattice. As Kernel Average Misorientation (KAM) maps can be derived from Electron Backscatter Diffraction (EBSD) scans, the question arose whether plastic zones could be located using KAM maps deduced from EBSD scans. To test this idea, the following experiment was designed. Flat aluminum alloy 6061-T4 specimens of 220 mm × 30 mm × 1 mm with 6 mm notches were cut out. Fatigue cracks at these notches were propagated under different loads at 30 Hz. KAM maps of the region around the tip were produced of three samples at grid sizes between 0.3 and 2 μm and maximum misorientation of 5° because larger ones are due to grain boundaries. The KAM maps however showed random misorientations everywhere and nothing specific was visible around the tip. Vickers micro-hardness measurements however did identify a plastic zone, which coincided with the theoretical calculations. A final KAM calibration test showed that a KAM measurement does not distinguish between noise and plastic strain lower than 2.2% strain. Because the plastic zone due to crack propagation produces less than this 2.2%, the KAM method was concluded to be inappropriate to identify the plastic zones caused by crack propagation.

Characteristics

The experiments in this category have the purpose to test a means recommended for achieving a pre-specified technical end; this purpose also defines their role in their hierarchical project plans. These experiments are often carried out at the end of a project. Typically, the prescribed action plan is followed and then the outcomes are assessed. Thus, the main independent variable is *nominal* (the prescribed order of actions) and the most proximate dependent variable is often *quantitative*. Sometimes the outcomes have intrinsic value, but normally the distant outcomes are more interesting. Thus, typically, the proximate outcomes are descriptive and the distant ones are normative.

Means-end knowledge is more general than a successful test of a working principle, i.e. a proof of concept. Consequently, the results of means-end knowledge tests are more comprehensive and less context dependent than those testing a working principle. The main difference between m-e knowledge and practical experiments is, therefore, that the first also includes comparisons of the consequences of alternative interventions whereas the second just measures whether a chosen working principle delivers what is expected. Note that for the outcome of a means-end tests to become real m-e knowledge, the most important side effects of all considered means should be considered as well.

At the end of the 19 century, William Froude observed that the friction from 6 ft to 12 ft models satisfied Frédéric Reech's scaling law, but the scaling from 3 ft to 6 ft did not. He hypothesized that to extrapolate the resistance of a large ship hull from that of a scale model, one needs to distinguish between the calculation of viscous friction and the scaling

of wave making resistance. Once Froude had obtained his viscous friction formula he combined it with Reech's scaling law to predict the complete friction of the *HMS Greyhound* and validated it with his famous 'Experiments for the Determination of the Resistance of a Full-sized Ship, at Various Speeds, by Trials with *HMS Greyhound*, Performed Off Portsmouth in August and September 1871'. Froude's predictions turned out to be approximately true and his experiments convinced most naval architects that Froude's scaling method was valid. The *Greyhound validation experiments* make the ideal type of experiments testing general m-e knowledge.

AlphaZero type – testing computer outcomes

The introduction and development of digital computers have had an enormous influence on techno-science as it increased computation and communication capacities beyond imagination. Today, large computation intensive models and simulations have become standard equipment for engineers and scientists. Many engineering experiments aim at establishing trust in simulations, algorithms, finite element models, or finding the appropriate simulation software.²⁵ The following is an experiment to see if a computer test can take the place of empirical tests of cartilage damage.

In the context of detecting cartilage damage in a human ankle, the question arose whether the program *Wave2000*[®] is accurate enough to describe real-life responses to ultrasonic tests. To answer this question, a test was set up with a simple Perspex model. The talus (a round disk) and tibia (beam with cutout fitting this disk at close distance) were connected to a frame with source and receiver transducers at either side of the joint space. An ultrasound pulse (1 MHz center frequency) was transmitted and received. The set-up was immersed in water and 2D-modelled in *Wave2000*[®]. Twenty test-retest experiments fell within the 95% reproducibility interval. Ten situations were considered: no damage and 2 mm deep damages on the disk of various widths (2, 4 and 6 mm) at the middle between the two transducers (90°), close to the source (120°), and close to the receiver (60°). A comparison between the computational simulations and the real measurements revealed a normalized cross-correlation less than the required 95% in six of the measurement situations. For all computational situations the pulse arrived before it arrived in reality. Moreover, in the experiment the waves did not flatten out after 6e-5s (perhaps due to reflections), whereas in the computational simulation they did. The researchers did not find trends in the ways the damage size and location affected the output signal. It was advised to use shorter wavelengths.

Characteristics

The experiments in this category have the purpose of testing or validating the outcomes of simulations, or other computational methods for specific applications. This purpose is normally straightforward and fixes the roles of these experiments in the embedding project plan. As the purpose of these experiments is validation of the mostly quantitative and descriptive (in)dependent variables, the value of these experiments typically equates to the instrumental value of the distant purposes. The experiments are context dependent as far as they test computational outcomes for specific circumstances, but the general purposes of these programs bestow them with some context independence.

In 2016, world-champion Go player, Lee Sedol, was beaten by AlphaGo, something that had been unimaginable before because of the complexity of GO. In 2018, Davis Silver and his coauthors introduced proposals about how to generalize the algorithms of AlphaGo, towards a generally applicable deep-learning AI program AlphaZero.²⁶ They tested it by letting it teach itself Chess, Shogi and Go without supervision, and made the resulting programs contest with the world champion programs Stockfish, Elmo and AlphaGo Zero. All three were defeated in astonishing ways. For this reason, experiments that test general computational methods for a specific application are put under the banner *AlphaZero tests*. The difference between the AlphaZero and the *Greyhound* experiments is that the latter tests whether the manipulation of reality indeed produces the required results, while the former tests the extent to which virtual reality constructed entirely from empirical laws as we believe them resembles simulated reality.

Trial-and-Error – preliminary type, establishing grip

Finally, experiments are carried out not only to ask nature or an artifact well-formed questions, let alone to test their answers, but also to help to formulate such well-formed questions.²⁷ These experiments are preliminary, and precede a design, descriptive or m-e knowledge project. Our database only contains a few of them. In design, preliminary investigations are typically carried out using simulation software. Our example concerns an experiment preceding m-e knowledge.

In the search for efficiently co-firing biomass in existing coal plants, grinding effects of torrefied biomass were researched. To that end, bagasse (TBG), roadside grass (TRSG) and wood chips (TWC) were torrefied for two hours, at 270° Celsius in a Nitrogen flow of 400L/hr. Next, three samples for each were grinded with a ball mill for 0.5, 1, 5, 10 and 15 min, and the temperature and particle size distributions were measured. To the surprise of the researchers the temperature during the grinding decreased approximately 2.5°C for all grinding times. They found this hard to explain. For all samples, the particle size no longer decreased after some time, which depended on the biomass. The differences in grindability were most obvious after 30s of milling when TBG had particles of the smallest size, and all other particles were smaller than 50 microns; TRSG and TWC showed a larger minimum particle size after 30s of milling. TBG was the best and TRSG was the hardest to grind. Its minimum particle size was only obtained after 10 min; moreover, it was less brittle than the other materials. TWC reached its minimum particle size after five minutes. It was shattered because one minute of milling produced particles of the smallest size, while chunks of material were present as well. The researchers were unable to quantify the influence of torrefaction of biomass. Nevertheless, according to observation with the naked eye, torrefaction significantly improved grindability of the biomasses. It was concluded that the research had to be continued with larger batches of biomass and an infrared thermometer; in addition, it was advised to investigate the relationship between torrefaction temperatures and grinding performance.

Characteristics

Preliminary experiments help to identify the relevant starting points for research projects. They help to get a grip on some phenomenon. In descriptive knowledge projects, their

function is mainly to identify the relevant variables or specify the subject to be investigated. For m-e knowledge projects, however, their job is to establish technical actions or interventions that will achieve the goal stated in the canonical project description. Because of the virtual absence of knowledge about the relevant variables, hypotheses, mechanisms let alone theories, the preliminary experiments are referred to as *Trial-and-Error Experiments*.

The theoretical frame

The typology expounded above bears similarities to two distinctions found in the recent literature. The first originates in the work of Friedrich Steinle²⁸ and Richard Burian,²⁹ and is the difference between *exploratory* and *theory-driven* experiments. The latter ‘are done with a well-formed theory in mind, from the very first idea, via the specific design and the execution, to the evaluation’.³⁰ ‘Exploratory experimentation, in contrast, is driven by elementary desire to obtain empirical regularities and to find out concepts and classifications by means of which those regularities can be formulated’.³¹ Note that to emphasize the action aspect of experiments Steinle prefers ‘experimentation’ to ‘experiment’. Moreover, underlining that not ‘all experimentation should be subsumed under these two types’ and ‘exploratory experimentation ... is not the counterpart of theory-driven experimentation’ Steinle acknowledges the incompleteness of his typology.³² For Steinle exploratory experimentation is epistemologically important because they lead to new conceptualizations. For instance, Faraday’s magnetic lines of force preceded the electromagnetic field theory. Steinle’s distinction subscribes to an order in time, since exploratory research precedes the tests of hypotheses. Thus, it is referred to here as the ‘horizontal’ distinction.

In contrast to Steinle’s focus on descriptive knowledge, Sven Ove Hansson introduces the distinction between *epistemic* and (directly) *action-guiding* experiments.³³ He writes:

An epistemic experiment aims at providing information about the workings of the world we live in. Therefore, the regularities looked for are such that can reveal mechanisms and propensities of the study objects.... In contrast, a directly action-guiding experiment has a practical purpose. It is performed to find out whether some intervention can be used to achieve a specified practical purpose. (Hansson 2019, p.2 of 23)

Hansson’s distinction, as that of Steinle, is based on difference in purpose. According to himself, Hansson actually distinguishes between ‘interpretations of experiments’ rather than experiments themselves,³⁴ because the same experiment (as protocol) can be interpreted as epistemic and action-guiding.³⁵ This difference between the experiment as protocol and its aim (or interpretation) has been addressed in section 2. Contrary to Steinle’s distinction, Hansson brings us outside the realm of descriptive knowledge and introduces appropriateness of interventions in technology, management, health care, and agriculture, etc. Clinical and field trials are typical examples. Hansson’s knowledge-intervention distinction will be referred to as a ‘vertical’ one.

If the horizontal distinction concerns the order in time, and the vertical is about types of outcomes, then the two are at right angles and independent. The 2×2 matrix provided by their combination already harbors four types presented above. Stress/strain-relationship experiments are of the exploring hypothesis type, whereas Boyle’s experiments are of the hypothesis testing type. Similarly, the *Swan/Raven* practical experiments are action explorative, (as the experiments of the Froude’s Torquay type), whereas the *Greyhound*

Table 2. The final classification.

Context	Testing	Determining	
<i>Descriptive knowledge</i> (general)	(1) BOYLE Hypothesis testing	(4) STRESS/STRAIN Value or Relation between variables with intrinsic value	(8) TRIAL & ERROR
<i>Designs</i> (unique working principle)	(2) WRIGHT FLYER Design testing	(5) EDISON FILAMENTS Working principle for unique design (proof concept)	
<i>M-e knowledge</i> (general)	(6) GREYHOUND m-e knowledge testing	(3) SWAN/RAVEN Optimal actions (practical experiment)	
		(4) TORQUAY TOWING TANK Value or Relation between variables with low intrinsic value	
Computational	(7) ALPHAZERO simulation, model algorithm testing		

experiments are of the action testing type. As Hansson does not distinguish between specific design and general means-end knowledge, a third row should be added to the matrix, which concerns the experiments in which appropriate working principle for a unique design are explored, the Edison filament finding type, and those in which designs are tested, the *Wright Flyer* testing type. The 2×3 matrix only misses the validation of computational output, and pure trial & error experiments, which precede almost all knowledge and experience. This results in the overview in Table 2.

Let us make some brief observations. First, the distinction between determining and testing does not coincide exactly with the one of Steinle. He writes for instance: ‘Theory-driven experimentation is not necessarily a test of theories or of hypotheses’.³⁶ Nevertheless, his exploratory experiments come close to our trial-and-error ones, which, regarding descriptive knowledge, often result in new concepts or classifications. Second, according to Hansson, epistemic and action-guiding experiments aim both at knowledge,³⁷ and he does not explicitly distinguish between descriptive and means-end knowledge, indispensable in our typology. Third, the ideal-type character of the present typology should be reemphasized. The boundaries between the types may sometimes not be as sharp as in a real classification. It might also be incomplete in the sense that perhaps it does not cover all laboratory experiments. Regarding the present database, however, the eight categories considered are the typical focus points of similarity and distinction.

Our final remark concerns the relationship between the present typology of experiments and that of engineering projects presented by Zwart and de Vries.³⁸ They distinguish between six project types based on differences in final goals and applied methods, viz., 1. descriptive knowledge about the world; 2. a designed artifact or processes; 3. general means-end knowledge, know how or guidelines formulating how to achieve some (technical) goal; 4. specific models or simulations; 5. optimizations; and 6. Mathematical and information science results. Although these types do not seem completely unrelated to the experiment types discerned, their relationship is far from one-to-one. Here, the distinction between determining and testing turned out to be the most fundamental one when ordering the experiments in practice. It is closely related to the difference between the intervention and evaluation phases so fundamental in the acquisition of experience.³⁹ Indeed, for the three main project categories, descriptive and m-e knowledge and design, determining and testing experiments were clearly present – although for m-e knowledge

determining the qualitative type has to be distinguished from the quantitative one.⁴⁰ Design distorts the scheme a bit more because it covers also (unique) working principles. Whereas experiments test designs, experimenters mainly search for working principles and proofs of concepts for designs. Besides these tests, no preliminary design experiments were encountered. Finally, regarding the three other project types (models, optimizations and mathematics) the database does only contain validation experiments.

Relevance and outlook

Let us finally turn to some issues for follow-up research and conclusions. We start with addressing the methodological and philosophical relevance of the present typology. Regarding the first we will consider two issues: the importance of a clear description of experimental purpose and the quality assessment of experiments.

Almost all experiments are imbedded in a hierarchical project plan designed to achieve some predefined goal. Not elaborating all technical details, these plans are partial,⁴¹ and the same holds for the embedded experiments. Reading project plans, one often comes across experiment descriptions that fail to mention the exact role of the experiment in the project or the experiment's precise aim. The present typology will be helpful in developing a clear idea of the experiment's function description and its role in the embedding project. This is the first methodological issue. Sometimes experiments serve multiple goals. For instance, the first Nereda[®] aerobic granular sludge experiments served at least two purposes: falsification of the existing biological explanation of granular sludge and a proof of principle of granular sludge waste water treatment plants, with a much smaller footprint than traditional ones.⁴² Even for Trial-and-Error experiments a description of the more distance purpose pays off.

Regarding quality assessment, Hansson discusses the following safeguards for action-guiding experiments: control experiments; parameter variation; outcome measurement; blinding; randomization; and statistical evaluation.⁴³ He adds a normative claim, 'the safeguards needed to avoid mistakes are essentially the same in action-guiding and epistemic experiments' and a descriptive one, 'these safeguards have to a large extent been developed for action-guiding experiments'.⁴⁴ The conjunction of both claims triggers the question why the safeguards did not develop in relation to epistemic experiments? The normative claim is an interesting one but needs qualification. We already saw different interpretations of parameter variations. But what is more important, stress/strain curve experiments are not double blind; nor were those of Edison, the Wright brothers or Froude. Double blind experiments are much more relevant in cases where people are tested.

Overall, controlled laboratory experiments are typically defined by an intervention, randomization, pre- and after measurements, and a control group. In many experiments described above the same measurements are repeated numerous times for reasons of (inter-laboratory) precision and accuracy. Experiments are even carried out to assess between-laboratory reproducibility. But if we test materials and not people, control group experiments are absent almost everywhere. Once the elasticity of some steel alloy is known, it is assumed to remain the same during the time another piece of the same material undergoes some treatment. Although hysteresis is important for many phenomena in

mechanics, the Solomon-four group design does not make sense for determining material properties. Thus, safeguards may vary with the subject of the experiments.

In 1935 Ronald Fisher published his *The Design of Experiments* (DoE), which paved the way for the application of modern statistics to experimentation. In chemical engineering for instance, blocking techniques and factorial analyses of variances are used to find effectively optimal circumstances to produce complex chemical compounds. Although Hansson correctly claims that safeguards are important for experiments, the exact relationships between various safeguards (and other statistical methods) on the one hand, and the various experimental types on the other make an interesting subject for further research. When the independent variables are nominal, such as the different configurations in the steering tip experiments (*Swan/Raven* example) no correlation with the quantitative dependent variables are calculated. Because some goals in techno-science only require 'good-enough results' (satisficing), it might turn out that the level of statistical rigor or the safeguards needed may vary with the type of experiments.⁴⁵ This is the second methodological issue yielded by our typology of experiments.

Philosophically the relevance of our typology stems mainly from its action-plan analysis.⁴⁶ Conceiving experiments as well-designed action plans reveals their position in the embedding project and prepares the way to research the various means-end hierarchies in experiments themselves. At first sight, this hierarchy might be more constrained than that of entire research projects where possibilities of combinations seem endless. When experimental physicists decide between different possibilities of illumination, which relates to values such as resolution, dissipated energy, heat production, etc., then that sub-decision of their experiment is a technical one. These decisions, however, do not make the entire experiment a technical one. In this sense, Boyle's choice to put his glass U-tube in a block of wood to prevent it from crashing under the high pressures due to the addition of mercury does not make his experiment a technical nor a practical one. Despite these technical aspects, the goal of his experiment was to come to a mathematical relation between pressure and volume.

The hierarchical means-end analysis of experiment actions also has philosophical impact, for instance regarding claims about theory-ladenness. It provides the opportunity to investigate the roles theories play in experimentation. Refraining from carrying out stress/strain experiments double blind or with control groups implies that some physical properties of steel alloys are assumed to be independent of time. Or, to mention another example, it will help to analyze the influence of optical geometry on the outcomes of experiments using optical instruments, based on this geometry. Analyzing the means-end hierarchy of experiments will help to discriminate between different types of theory-ladenness and to find out the extent to which it is problematic.

More generally, a typology of experiments can put the debates about truth and (anti)realism in a new perspective. For instance, William James wrote: 'The true ... is only the expedient in the way of our thinking, just as the right is only the expedient in the way of our behaving'.⁴⁷ If truth is reduced to being expedient, it becomes difficult to distinguish between the Boyle and the *Swan/Raven* type of experiments; the second is practical whereas the first is not (despite the possible application of its results). Another interesting issue is to what extent all experimentation is equally guilty of Richard Rorty's stumbling block: 'representationalism'.⁴⁸ Perhaps practical experiments are less representationalist than Boyle-type of experiments, and might form the basis of pragmatism. Similarly, we

may elaborate Ian Hacking's (1983) shift from the realism-antirealism debate toward representing and intervening. To that purpose, we need to investigate the different ways in which experiments and technologies intervene and come to a more balanced account of the relationship between representations and interventions in techno-science.

Hopefully classifying the types of engineering experiments not only sheds new light on the delicate hierarchical means-end networks in techno-science research projects at various levels, but also opens up a broader philosophical interest in the foundations of engineering and technology per se.

Notes

1. Marcorini, *The History of Science and Technology*, 367.
2. Gordon, "10 Moments That Made American Business."
3. Franklin and Perovic, "Experiment in Physics."
4. Hacking, *Representing and Intervening*.
5. Cf. Vaclav Smil's characterization of the period 1867–1914 as 'The Age of Synergy'. Smil, *Creating the Twentieth Century*.
6. Zwart and de Vries, "Methodological Classification of Innovative Engineering Projects."
7. We take engineering projects as plans in the sense of Bratman's Intention, plans, and practical reason.
8. Weber, "Objectivity in Social Science", 90.
9. Norton, *Dynamic Fields and Waves*, 83.
10. Shankland, "Michelson–Morley Experiment," 32.
11. For similar reasons Hansson writes "we should distinguish between ... interpretations of experiments. It is conceivable for one and the same experiment to be used for both purposes, i.e. interpreted in both ways". Hansson, "Farmers' Experiments and Scientific Methodology," note 2.
12. Bratman, *Intention, Plans, and Practical Reason*, section 3.1.
13. Einstein: "If the Michelson–Morley experiment had not brought us into serious embarrassment, no one would have regarded the relativity theory as a (halfway) redemption." Fölsing, *Albert Einstein*.
14. Note that this distinction depends on thick descriptions.
15. Boyle, A defence of the doctrine touching the spring and weight of the air, 58. To what extent the historical Boyle was ambivalent in his commitment to "Boyle type" experiments we leave to the historians of science. Here it should be taken as just a label.
16. Hansson, "Experiments Before Science."
17. In practice, developing the details of a design, and testing them, often alternate in an iterative design process.
18. Hansson, "Farmers' Experiments and Scientific Methodology." Kroes's practical experiments also seem similar, but the learning involved there is suggested not to be based on "regularities and control" but on storytelling. Kroes, "Design Methodology and the Nature of Technical Artefacts," 32.
19. Here, we leave aside that the Swan/Raven experiments inspired Froude to research the question why the 6ft to 12ft models scaling satisfied Frédéric Reech's (1805–1884) scaling law, and the scaling from 3ft to 6ft did not. This led eventually to Froude's famous quantitative scaling method.
20. Vincenti, *What Engineers Know and How They Know it*, 139.
21. E.g. Carlton, *Marine Propellers and Propulsion*, 93.
22. Steinle, "Experiments in History and Philosophy of Science", S70.
23. Hansson, "Experiments Before Science," 99.
24. Cf. e.g. Stanley and Williamson, "Knowing How."
25. This does not make the computational simulations themselves physical experiments.
26. Silver et al., "A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go Through Self-Play."

27. This type also harbors those experiments carried out outside a specific research project living “a life of [their] own.” Hacking, *Representing and Intervening*, xiii; Franklin and Perovic, “Experiment in Physics”.
28. Steinle “Experiments in History and Philosophy of Science.”
29. Burian, “Exploratory Experimentation”.
30. Steinle, “Experiments in History and Philosophy of Science,” S69.
31. Steinle, *Ibid.*, S70. See also Franklin, “Exploratory Experiments”; Elliott, “Varieties of Exploratory Experimentation in Nanotoxicology”; Waters, “The Nature and Context of Exploratory Experimentation”; Karaca, “A Case Study in Experimental Exploration.”
32. Steinle, “Experiments in History,” (S69) and (S71), respectively.
33. Hansson, “Experiments Before Science.”; “Experiments: Why and How?”; “Farmers’ Experiments and Scientific Methodology.”
34. Hansson, “Farmers’ Experiments and Scientific Methodology,” note 2.
35. Hansson, “Experiments Before Science,” 92.
36. Steinle, “Entering New Fields,” S69.
37. Hansson, “Farmers’ Experiments and Scientific Methodology,” note 2,
38. Zwart and de Vries, “Methodological Classification of Innovative Engineering Projects.”
39. De Groot, *Methodology*, section 1.1.
40. Note that the use of the six types mentioned needs the criteria 3–5 of section 2.
41. Bratman, *Intention, Plans, and Practical Reason*, sect.3.1.
42. Zwart and Kroes, “Substantive and Procedural Contexts of Engineering Design.”
43. Hansson, “Experiments Before Science,” 99.
44. Hansson, *Ibid.*, 99.
45. Maarten Franssen pointed out this issue.
46. It already gave rise to the acknowledgment of the directionless experiments in section 3.1.1.
47. James, *The Meaning of Truth*, preface.
48. Rorty, *Philosophy and the Mirror of Nature*. “I argue that the attempt (which has defined traditional philosophy) to explicate “rationality” and “objectivity” in terms of conditions of accurate representation is a self-deceptive effort to eternalize the normal discourse of the day” p.11.

Acknowledgements

The author acknowledges fruitful discussions about experiments with Maarten Franssen and Léna Soler.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Bertram, V. *Practical Ship Hydrodynamics*. Amsterdam/Boston/Paris: Elsevier, 2012.
- Boyle, R. *New Experimental: Physico-Mechanical, Touching the Spring of the air, and its Effects; Made, for the Most Part, in a new Pneumatical Engine*. Written by the way of Letter, to the right honorable Charles, Lord Viscount of Dungarvan, eldest son to the Earle of Corke. Oxford: H. Hall, printer to the University, 1660.
- Boyle, R. *A Defence of the Doctrine Touching the Spring and Weight of the air Propos’d by Mr. R. Boyle in his New Physico-Mechanical Experiments, Against the Objections of Franciscus Linus*. London: Richard Davis, 1662.
- Bratman, M. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press, 1987.
- Burian, Richard M. “Exploratory Experimentation and the Role of Histochemical Techniques in the Work of Jean Brachet, 1938-1952.” *History and Philosophy of the Life Sciences* 19, no. 1 (1997): 27–45.
- Carlton, J. *Marine Propellers and Propulsion*. 3rd ed. Amsterdam: Butterworth-Heinemann, 2012.

- Elliott, K. C. "Varieties of Exploratory Experimentation in Nanotoxicology." *History and Philosophy of the Life Sciences* 29, no. 3 (2007): 313–336.
- Fisher, R. A. *The Design of Experiments*. New York: Macmillan Pub Co, 1935.
- Fölsing, A. *Albert Einstein: A Biography*. New York: Penguin Group, 1998.
- Franklin, L. R. "Exploratory Experiments." *Philosophy of Science* 72, no. 5 (2005): 888–899.
- Franklin, A., and S. Perovic. "Experiment in Physics." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta, Winter 2019. <https://plato.stanford.edu/archives/win2019/entries/physics-experiment/>.
- Froude, W. "On Experiments with HMS Greyhound." *Transactions of the Royal Institution of Naval Architects* 15 (1874): 36–73.
- Gordon, J. S. "10 Moments That Made American Business." *American Heritage*, February/March, 2007.
- Groot, A. D. de. *Methodology; Foundations of Inference and Research in the Behavioral Sciences*. The Hague: Mouton, 1969.
- Hacking, I. *Representing and Intervening*. Cambridge: Cambridge University Press, 1983.
- Hansson, S. O. "Experiments Before Science. What Science Learned from Technological Experiments." In *The Role of Technology in Science: Philosophical Perspectives*, edited by S. O. Hansson, 81–110, 2015. doi:10.1007/978-94-017-9762-7_5.
- Hansson, S. O. "Experiments: Why and How?" *Science and Engineering Ethics* 22, no. 3 (2016): 613–632. doi:10.1007/s11948-015-9635-3.
- Hansson, S. O. "Farmers' Experiments and Scientific Methodology." *European Journal for Philosophy of Science* 9, no. 3 (2019): 1–23.
- James, W. *The Meaning of Truth: A Sequel to Pragmatism*. New York: Longman, Greens, 1909.
- Karaca, K. "A Case Study in Experimental Exploration: Exploratory Data Selection at the Large Hadron Collider." *Synthese* 194, no. 2 (2017): 333–354.
- Kroes, P. "Design Methodology and the Nature of Technical Artefacts." *Design Studies, Philosophy of Design* 23, no. 3 (2002): 287–302.
- Kroes, P. "Control in Scientific and Practical Experiments." In *New Perspectives on Technology in Society*, edited by I. R. van de Poel, L. Asveld, and D. C. Mehos, 16–35. Abingdon: Routledge, 2017.
- Li, R., C. Zhong, Y. Yu, H. Liu, M. Sakurai, L. Yu, ... J. C. Izpisua Belmonte. "Generation of Blastocyst-Like Structures from Mouse Embryonic and Adult Cell Cultures." *Cell* 179, no. 3 (2019): 687–702.
- Marcorini, E. *The History of Science and Technology: A Narrative Chronology*. New York: Facts on file, 1988.
- Niu, Y., N. Sun, C. Li, Y. Lei, Z. Huang, J. Wu, T. Tan. "Dissecting Primate Early Post-Implantation Development Using Long-Term in Vitro Embryo Culture." *Science* 366, no. 6467 (2019): eaaw5754.
- Norton, A. *Dynamic Fields and Waves*. Boca Raton: CRC Press, 2000.
- Peterson, D. A., G. C. Littlewort, ... T. J. Sejnowski. "Objective, Computerized Video-Based Rating of Blepharospasm Severity." *Neurology* 87, no. 20 (2016): 2146–2153.
- Rorty, R. *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press, 1979.
- Shankland, R. S. "Michelson–Morley Experiment." *American Journal of Physics* 31, no. 1 (1964): 16–35.
- Silver, D., T. Hubert, J. Schrittwieser, ... D. Hassabis. "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go Through Self-Play." *Science* 362, no. 6419 (2018): 1140–1144.
- Smil, V. *Creating the Twentieth Century: Technical Innovations of 1867–1914 and Their Lasting Impact*. Oxford/New York: Oxford University Press, 2005.
- Stanley, J., and T. Williamson. "Knowing How." *The Journal of Philosophy* 98, no. 8 (2001): 411–444.
- Steinle, F. "Entering New Fields: Exploratory Uses of Experimentation." *Philosophy of Science* 64 (1997): S65–S74.
- Steinle, F. "Experiments in History and Philosophy of Science." *Perspectives on Science* 10 (2002): 408–432.
- Vincenti, W. *What Engineers Know and how They Know it: Analytical Studies from Aeronautical History*. Baltimore: Johns Hopkins University Press, 1990.
- Waters, C. Kenneth. "The Nature and Context of Exploratory Experimentation: An Introduction to Three Case Studies of Exploratory Research." *History and Philosophy of the Life Sciences* 29, no. 3 (2007): 275–284.

- Weber, Max. "Objectivity in Social Science and Social Policy." In *The Methodology of the Social Sciences*, edited by E. A. Shils, and H. A. Finch, 49–112. New York: Free Press, [1949](#).
- Zwart, S. D., and P. Kroes. "Substantive and Procedural Contexts of Engineering Design." In *Engineering Identities, Epistemologies and Values*, edited by S. H. Christensen, C. Didier, A. Jamison, M. Meganck, C. Mitcham, and B. Newberry, 381–400. Cham: Springer, [2015](#).
- Zwart, S. D., and M. J. de Vries. "Methodological Classification of Innovative Engineering Projects." In *Philosophy of Technology After the Empirical Turn*, edited by M. Franssen, P. E. Vermaas, P. Kroes, and A. W. M. Meijers, 219–248. Cham: Springer, [2016](#).