A short history
of the method of thought experiments
in science

Ulrich Kühne

KUHNE@UNI-BREMEN.DE

1) Remarks on the historical development of the *experimental* method

2) Major stages in the development of a method called "*thought experiment*"

- **(a) Metaphysics (Kant, Ørsted)**
- **(b) Anti-Metaphysics (Mach)**
- **(c) The modern physicists (Einstein, Heisenberg)**

3) Does the history of the concept allow for a meaningful explanation of "thought experiments"?

- **(a) Proposal 1: The construction of monsters (Lakatos)**
- **(b) Proposal 2: Arguments using generalised principles**

Literature:

Ulrich Kühne: *Die Methode des Gedankenexperiments*. Suhrkamp Verlag, Frankfurt am Main (Germany) 2005. 410 pages, 14 Euro (≈ 21 Can. $)

1) Remarks on the historical development of the *experimental* method

Some people claim, the “method of thought experiments” started with the pre-Socratics. This implies that the method of thought experiments is just what people do when they start to think about nature, asking “What if?” and speculating about alternative creations of our world. – We all know the wonderful examples of philosophical riddles in the tradition of the Eleatic School like Achilles and the Tortoise and many others.

On the other hand: It doesn’t really make sense to call these beginnings of philosophy (not: science!) “thought experiments”, since at this time the experimental method did not exist. Of course, at the beginning of all science there is observation and speculation on how to explain what is observed, i.e. an intellectual activity. But the experimental method – and thus the thought-experimental method – is something quite different from that.

The experimental method is a quite recent invention by some truly revolutionary thinkers round about 1600 AD, particularly Simon Stevin, William Gilbert, Francis Bacon and Galileo Galilei. It’s an invention, which was put forward against very strong opposition from philosophers and against the obvious common sense of the common people. It is so revolutionary that even today, despite the spectacular successes of experimental science during the past 400 years, it is still not generally accepted by the common people nor by philosophers.

An essential feature of the experimental method is that the scientist no longer looks for a “true” explanation of nature – true in the Aristotelian sense of “theory depicting the observed phenomenon”, the “mapping” of phenomena facts to theory-sentences.

For instance: it is “true” that under normal circumstances bodies of different weight fall (slightly) differently. The “Aristotelian Law of Free Fall” had an intuitively plausible explanation for this fact.

Contrary to this, modern science does not attempt to **picture** nature in her theories, but to **reconstruct** nature. Reconstruction means that the scientist tries to rebuild the observed phenomenon out of *idealized singular factors* as building blocks, fully knowing that each of these idealized singular factors does not appear anywhere in nature and can in principle not be observed and is therefore in an Aristotelian sense “false”.

Rather than a passive observation of nature the experiment is an active interference with nature. The experiment tries to create by approximation an artificially produced so-called “pure state”. To save the phenomenon many of these singular factors have to be **superimposed** to a common joint outcome.

Such singular factors are for instance the law of inertia, the law of free fall (frictionless in vacuum), etc. Such laws are not only counterintuitive. They are also counterfactual (i.e. false), because there is nowhere on Earth (outside a human made laboratory) a frictionless motion of inertia or a true vacuum.

Aristotle, who was at the core of his philosophy interested in explaining the real, obvious and openly visible phenomena of nature, would have called all such idealised natural laws at best useless fictional fantasies, if not utterly absurd. He not only “would have”, he actually did: cf. his *Physica,* book *Delta*.

Albert Einstein pointed out this difference very clearly in his popular science book *The Evolution of Physics* (co-authored with Leopold Infeld): “In a good mystery story the most obvious clues often lead to the wrong suspects. In our attempts to understand the laws of nature we find, similarly, that the most obvious intuitive explanation is often the wrong one. Human thought creates an ever-changing picture of the universe. Galileo’s contribution was to destroy the intuitive view and replace it by a new one. This is the significance of Galileo's discovery."[1]

(Foreshadowing: This of course is fundamental to our understanding of thought experiments! Galileo and Einstein should not be promulgated as inventors of intuitive thought-experiments, if they so loudly claim to be the leaders of the great philosophical revolution to overthrow and banish all intuitive thinking from science.)

What is this "new view", which Galileo (according to Einstein) invented to replace intuitive thinking? Reconstruction! By superimposing many singular factors – for instance (1) inertia *plus* (2) gravity *plus* (3) friction *plus* (4) buoyancy – the visible phenomena of nature can be reconstructed. And, in the superposition of many of those factors, the reconstruction is far more successful and more precise than any previous attempt to find a single phenomenological law in the "direct" way, skipping the deviation over the idealized counterfactual single-issue laws.

The most prominent example of the reconstruction of nature by superposition of singular factors is the famous "parallelogram of forces" which was invented by Stevin and Galileo in several papers between 1586 and 1638.

Having this in mind, one can easily see that the original version of Galileo's famous "thought experiment on the free fall" (first chapter of the *Discorsi* 1638) can easily be recognised as a persiflage on the pre-modern understanding of science by Galileo's contemporary followers of the Aristotelian philosophy. Galileo (in the words of Salviati) succeeds in confusing the seemingly most obvious intuitions – such that at the end the Aristotelian Simplicio does not trust his intuitions at all any more. The Aristotelian method of science is beaten by its own weapons. The question, however, what the true law of free fall looks like is not and is not intended to be answered by this thought experiment.

[1] Chapter „I. The rise of the mechanical view", section 2: „The first clue"

That is obvious, because one half of the answer is known to the three characters well before they discuss the so called thought experiment proper: all three agree that they pretty well know how bodies of different weight from different material and from different heights fall *in reality* – namely with nearly the same speed.

The other half of the answer, namely Galileo's invention of the revolutionary new mathematical law of free fall, which takes into account gravity, buoyancy, and friction and which predicts a time-dependent function of instantaneous velocity in numerical quantities – all this is done in the later chapters of the *Discorsi* far later than the so-called thought experiment from the first chapter (which we should call "an introduction"). It is done by carefully considering empirical data and using a richly filled toolbox of mathematical methods and arguments, not by thought experiments.

2) Major stages in the development of a method called "*thought experiment*"

(a) Metaphysics (Kant, Ørsted)

The idea that the empirical side of science could at least partially be replaced by a "thought experiment", re-emerged much later into science – after this idea, which was common from antiquity till Galileo, was nearly extinct by the overwhelming success of the Galilean method of experimental science. In the first centuries after Galileo this idea was rather ridiculed. Enlightened people pictured it as the method of obscurantist monks of the medieval dark age in their futile efforts to decode the secrets of the universe from inside a monastic study without windows to the outside world, just using old scriptures and "reason".

The career of a newly invented scientific method called “thought experiment” started with the German philosopher Immanuel Kant. In fact, his plan was just opposite to what he accomplished. His actual plan was to **criticise** excessive and untenable claims of reason. To criticise a concept of reason, which is tightly linked to a long and futile tradition of metaphysics and to replace it with a concept of reason, which Kant explicitly took from the natural sciences. Natural science, to Kant, was the paradigm and paragon of the proper use of reason – and should thus be copied in philosophy.

That is: Kant’s intention was **not** the other way round: to improve scientific methodology from something he discovered in philosophy. (For that reason we find the word “critique” in the title of all three of his main books.)

But in an attempt to draw the exact line between the mere form of natural laws, which Kant sees in fact determined by reason, and the material content of natural laws, which can only be determined by experience, Kant wrote immediately after his *Critique of Pure Reason* a little book on physics, the *Metaphysical Foundations of Natural Science* (*Metaphysische Anfangsgründe der Naturwissenschaft*, 1786), which caused a dramatic development.

Firstly the book earned the great applause of a whole generation. What Kant did in this book was to reconstruct Newtonian physics in such a way that it seems clear that the core of Newtonian physics can be understood by reason alone – that is: he tried to debunk the purely formal structure in the mechanical laws of nature.

But Kant was understood by his followers and successors to have proven that a slightly more intelligent person than Newton himself could have come up with Newtonian physics without any empirical import. And if this worked so well with Newtonian physics, we should be optimistic that some true genius would soon find out all the other things about nature by mere philosophical speculation. This was the beginning of so-called "German Idealism" with such prominent members as Schelling and Hegel.

In this ambient the Danish physicist and philosopher Hans Christian Ørsted coined the term "thought experiment" first in 1811. His intention was to unite his adoration for Kantian metaphysics with traditional ideas about the hypothetical-deductive method of science. Compared with most other friends of German Idealism he was a very careful Kantian and quite sceptical towards the excessive ambitions and claims of other followers of that school.

What followed is a big tragic irony in the history of philosophy. It was just this friend of German Idealism Hans Christian Ørsted, who (involuntarily and never admitting it) destroyed it – by a real experiment not a thought experiment:

One central a-priori-truth from the *Metaphysical Foundations of Natural Science*, which Kant derived from reason alone, was: *Forces act on straight lines only by either attraction or repulsion.* Obviously true to anyone who understands the concept of force and has a sober view on the very idea of natural law. Everything else would be just crazy.

However in 1819/20 Ørsted discovered electromagnetism: The effect of an electric current is a rotary-movement of a magnetic needle nearby. The magnetic needle is neither attracted nor repulsed, but turns perpendicular to the wire.

The insight that the electromagnetic force has no mechanical model and thus cannot be reconstructed by forces acting on straight lines started some of the most revolutionary developments in physics and philosophy of the 19th century – but it killed all hopes to contribute to the progress of science by employing Kantian-like modes of thinking. Nature just **is** crazy, and no philosophy (which almost by definition tries to bring order and reason into our understanding of nature) can cope with this fact.

(b) Anti-Metaphysics (Mach)

The “method of thought experiments”, however, several times in later years rose from the grave (and died again). The next resurrection of the concept of thought experiment was again due to a progress of physics, which (again) looked like a triumph of rational insight over empirical confusion: the discovery of the theory of thermodynamics.

Around the middle of the 19^{th} century a new universally valid principle was promulgated by various authors: “Universal Energy Conservation”. In its easiest wording it says: “It is impossible by any means to build a perpetuum mobile.” This principle has a funny history, which we unfortunately cannot indulge in here. Just so much:

Despite the obvious fact that this principle was established against very forceful opposition from the side of philosophers and engineers (and against the seemingly most obvious empirical ‘facts’ like the eternal movements of the sky, the eternal movements of the sea, the wind and all animals), ...

... and despite the historical fact that this principle was derived from very, very careful generalisations from many singular laws of nature, all constantly scrutinized by hard experimental work, i.e. singular laws on (1) static (2) dynamics, (3) gravitation, (4) electromagnetism ... on (n) thermodynamics, (n+1) chemical combustion, and finally (n+2) animal nutrition and labour ...

... despite all this, soon after Joules, Meyer and Helmholtz formulated this principle in a universally generalized fashion, it suddenly sounded utterly familiar and intuitive to all scientists and philosophers.

And, again (as in the case of Newtonian physics and Kant), philosophers came up with the idea that one should be able to derive that "there is no perpetuum mobile" from philosophical reasons alone.

But worse this time: Scientists (not: philosophers) suddenly came up with arguments, which seem to prove this principle "the other way around": Just assuming this principle and seemingly nothing else, these scientist could perform the trick of deriving all of thermodynamics – all natural laws about heat, entropy, the ideal gas without even looking at a single thermometer.

It started with Sadi Carnot (1796-1832) and his famous Carnot-cycle (1824) and was completed in a beautifully closed mathematical theory by Josiah Willard Gibbs in his *Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundations of Thermodynamics* (1902). Einstein later emphasised that exactly this derivation of thermodynamics from the single, simple principle "there is no perpetuum mobile" was the central paradigm in his own discovery of relativity. Just one of very many of quotes of Einstein to this respect: "The method of relativity is in most parts analogue to thermodynamics, because thermodynamics is nothing else but a systematic answer to the question: How must the laws of nature be in order to make it impossible to build a perpetuum mobile?"[2]

I can't go here into the flaws of Einstein's view, however before him the Austrian experimental physicist Ernst Mach (1838-1916) had severe doubts about the seemingly overwhelming success of pure reason when it comes to thermodynamics.

Mach was a confessing anti-metaphysiscist and a radical opponent of Kant's philosophy. But, though not in the field of electromagnetism – Kant's research-project of the *Metaphysical Foundations of Natural Science* seemingly

[2] Albert Einstein: *Letters to Solovine*. New York 1987, p. 32f

succeeded in the field of thermodynamics. So Mach started a very fierce defence battle starting with his very first philosophical paper on the "History and Root of the Principle of the Conservation of Energy" (*Die Geschichte und die Wurzel des Satzes von der Erhaltung der Arbeit*, 1872 – long before his *Science of Mechanics* from 1883!), continuing till after his last (in his lifetime) published book on *Knowledge and Error* (1905) – and having (as we can see) nearly reached victory he gave up.

Victory would have been to show that Einstein was victim to a self-deception: To show that it is not possible to derive thermodynamics from principles. The derivation depends on background assumptions, which are contingent on different intellectual frameworks and incorporate empirical data, if the result turns out true. Some sort of a Duhem-Quine-thesis would do. Mach successfully proved this in his *Mechanics* with respect to the Archimedean law of the beam balance, but not with respect to thermodynamics. With thermodynamics this would be very tricky, since even Einstein failed to see this during the whole of his life.

But one reason, why Mach gave up so quickly, was that quite early in his philosophical career he invented something like an easy escape route on how to save the philosophy of anti-metaphysics despite strong evidence against it especially from thermodynamics. The trick was simply marketing: to re-interpret the success of reason in science by re-branding "reason" into "instinct".

"Instinct", of course, is the result of an adaptation of thinking to countless empirical data, which we humans collected during the million of years of our phylogenesis and the decades of our individual ontogenesis – before we can become fully fledged experts on metaphysics. In this empirical data from our phylo- and ontogenesis Mach discovered the true reason, why we so often find the "right guess" even in questions of advanced thermodynamics. No magic involved. Case closed.

However, it should be noted, that not even all of his friends where convinced by this evolutionary-biological camouflage of Kantian transcendental philosophy. Pierre Duhem, for instance did not. He frankly called all this thought-experimental stuff a great fake and a horrible example of logical confusion. Thought experiment transform physics quote “into a monstrous gibberish and aggregate just one *petitio principii* after the other and one *circulus vitiosus* after the other. A clearly thinking mind will refuse their permanent violation of logic with utter disgust ...”[3]

Strangely enough, Duhem in fact silenced Mach in this way on the topic of thought experiments. After Duhem’s *Aim and Structure of Physical Theory* (1906) Mach stopped writing on thought experiments. There is no way to declare yourself a strict opponent of metaphysics and favour “thought experiments” as a king’s road to scientific discovery.

(c) The modern physicists (Einstein, Heisenberg)

Today’s popularity of thought experiments obviously comes from Einstein. According to folk-history-of-science Einstein made his most prominent discoveries solely by use of thought experiments. The most famous thought experiments of Einstein are the train-TE (from which one can seemingly derive the Lorenz- transformation of special relativity), the lift-TE (by which he demonstrated the equivalence-principle of general relativity), or e.g. the photon-box-TE (which gives a simple derivation of the $E=mc^2$-formula).

However, one should be careful making such claims. Despite Einstein’s understanding of the logical origin of thermodynamics (which fits very well into such a picture), in all his works he was very reluctant with thought

[3] Pierre Duhem *La théorie physique, son objet et sa structure* (1906), translated from the reprint of the German edition of 1908 by Friedrich Adler, Hamburg: Meiner Verlag 1978, page 274.

experiment; sometimes openly **against** the use of thought experiments in science.

This opposition to the method of thought experiments had an obvious reason: The opponents of Einstein were the people, who used so-called "thought experiments" as their most powerful weapon to attack relativity. The most prominent leader of this group was Philipp Lenard, who started his career as a respected experimental scientist and assistant to Heinrich Hertz. He discovered the photoelectric effect which was the starting point of Einstein's Nobel-prize winning paper of 1905 on the quantum-interpretation of radiation. In fact Lenard won the Nobel-prize of 1905. But from around 1910 he turned into an opponent of the theory of relativity.

Lenard claimed that relativity contradicts the "plain and sane human understanding" ("gesunden, einfachen Menschenverstand")[4]. To prove this, he constructed thought experiments from simple scenarios like that of a train braking severely near a church-tower. While everything inside the train is shaken around, the church-tower suffers not the slightest effect. Therefore, Lenard claimed, relativity is wrong, since it simply isn't the case that we can transform this scenario into a different perspective, were the church-tower collapses and everything inside the train stays calm.[5]

Obviously, this is a wrong understanding of relativity. Einstein normally answered to this kind of objection that theory forming in modern physics follows utterly different criteria then to please the "plain and sane human

---

[4] Philipp Lenard: *passim*. E.g. Introduction to *Über Relativitätsprinzip, Äther, Gravitation* (first published 1917): „[...] und wir legen Wert darauf, hier ganz den Standpunkt des einfachen Verstands einzunehmen. [...] dieser einfache oder auch gesund zu nennende Verstand [...]"

[5] cf. *Über Relativitätsprinzip, Äther und Gravitation* in: Philipp Lenard, *Wissenschaftliche Abhandlungen*, *Band 4*. Posthumously edited by Charlotte Schönbeck, Diepholz/Berlin: GNT-Verlag 2003. (page 432 ff.) – Review of this book: Ulrich Kühne, „Ganz aus Galle und Intrigue: Die gesammelten Abhandlungen des nationalsozialistischen Physikers Philipp Lenard", *Süddeutsche Zeitung*, 22nd September 2003, p. 12.

understanding" – see for instance the quote above from Einstein's book *The Evolution of Physics*.[6]

Lenard was not convinced. From 1922, after he lost a public debate with Einstein on his thought experiments in Bad Nauheim, he turned openly anti-Semitic. In 1936/37 he published *Deutsche Physik*, the defining textbook in four volumes on the Nazi philosophy of science.[7] In the core of it he proposes the use of thought experiments as the methodology to find *true* theories – *true* meaning: being successfully tested against the plain and sane intuitions of the aryan bonehead. In his opinion members of the aryan ethnic are of a more practical and straightforward mode of thinking and are thus immune against any confusion by advanced theoretical physics or mathematical abstractions, which they by nature are unable to understand anyway. Strangely he thought this was to their advantage.

This conflict between abstract theories of mathematical physics and our simple intuitions – which the intuitions had actually lost (at the latest) in 1820 by Ørsted's discovery of electromagnetism – surprisingly started a new revival in the field of quantum mechanics. In his first paper on the uncertainty principle 1927 Heisenberg expressed his hope to derive quantum theory from an intuitive understanding of the measurement-process. To this end he invented several "thought experiments" in this paper, most prominently his "gamma-ray-microscope".

(Heisenberg at some times had some sympathies with philosophy of science in the Lenard-style, but lost these sympathies completely when in 1939 the supporters of Lenard prevented him becoming the successor of the chair of

[6] see also: Albert Einstein (1918): „Dialog über Einwände gegen die Relativitätstheorie" [Dialog on objections against the theory of relativity] *Die Naturwissenschaften*, No. 6, p. 697-702. (This was Einstein's direct answer to Lenard's „Über Relativitätsprinzip, Äther, Graviation" of 1917.)

[7] Philipp Lenard: *Deutsche Physik*, Munich: Lehman Verlag, volume 1 &2: 1936, 3 & 4: 1937. 2nd edition 1938, 3rd edition 1942. Scientific methodology is dealt with mainly in the general introduction in volume 1.

Arnold Sommerfeld in Munich on grounds that he was still too theoretical in his physics.)

Soon after Heisenberg finished this paper it became clear that there simply is no way to understand quantum-mechanical indeterminacies as resulting from disturbances caused by classical measurement processes – "no way" due to very convincing fundamental considerations.

Anyway, the conflict between quantum theory and traditional philosophical intuitions cannot be solved by any simple, plain and sane insight. There are many prominent examples, which probe different elaborate intuitions against quantum theory – Einstein-Podolsky-Rosen, Bell, Kochen-Specker –, but in all cases it was the intuitions, which had to change under the pressure of these arguments (or "thought experiments" if one likes the term).

3) Does the history of the concept allow for a meaningful explanation of "thought experiments"?

This very short historical overview might look quite depressing. We have seen that the so-called "thought experiment" has a strong tradition in the field of *pre-scientific* and even *anti-scientific* methodology. In this tradition "thought experiments" are explained in terms of *intuition* – the intuitions of children or philosophers (in the antiquity) before they start with real science. Or the intuitions of people, who were demanded too much by either the complexity and abstractness of advanced physical theories or by the craziness of nature.

But, besides this, I think, there is another tradition of thought experiments (not always called "thought experiments", like with Einstein), which clearly is *scientific*. In fact, I find two non-trivial elaborations on the concept of thought experiment, which

a) do have a history of successful application in science

b) may, with some reason, be called “thought experiments”
c) are, in principle, compatible with a general empirical picture of science, i.e.:
d) do not go into the business of unsubstantiated metaphysics.

(For being “trivial” I reject all explanations of thought experiments, which describe valid, well established methods in science, which we can perfectly well continue to call by their common names like “calculation”, “computer-simulation”, “use of arguments” etc.)

(a) Proposal 1: The construction of monsters (Lakatos)

The modern discussion of this proposal is normally centred around a rough idea by Thomas Kuhn, which has been elaborated and modified by Imre Lakatos in his *Proofs and Refutations* (1961/1976), a classic book on the methodology of mathematics.

In fact, this proposal originates not from natural science but from the question, how to do justice to the astonishing creativity in mathematics. The core idea of the answer again goes back to Kant, who claimed that mathematical reasoning is in fact not always logical deduction. As mere logical deduction, mathematics would be plain trivial. The steps, where according to Kant, the creative new ideas come into mathematics, he called “construction in intuition” (Anschauung). Thus, Kant claimed, mathematics is both, a priori true *and* informative.

Kuhn and Lakatos elaborated this idea by pointing out that even definitions of concepts have a quasi-empirical content build into them. This quasi-empirical content is the scope of the intended applications of the concept. In order to learn what counts as an intended application, one can construct imaginary scenarios and then decide, if in this scenario the concept still can be applied or not. The construction of these scenarios one can in a tradition of many philosophers call “thought experiments”.

However, in opposition to many others, it must be emphasised that one can **not** learn anything of interest from such imaginary applications of concepts, if we apply the concepts in a common or familiar fashion – if we apply the concepts in such a boring way that we can answer the question, if or if not the concept still can be applied, by just looking into our plain and sane intuitions.

The only possibility to learn something interesting from such scenarios arises if we construct, what Lakatos called a "monster", that is: a (nearly) nasty and intentional misunderstanding of the original definition of the concept in an utterly counter-intuitive extreme.

A nice example is the concept of "simultaneity", if looked upon in a scenario, where the "simultaneous" events under consideration take place at different ends of a train travelling at the speed of light. Suddenly one recognizes that the original intentions with this concept did contain ambivalences, which we would never have noticed, if we had not left the field of boring normal applications.

Though very fruitful this way, it is immediately clear that there is no way to determine in advance, how to reform or extend the meaning of a concept, after having discovered the semantic indeterminacy by this extreme-scenario. This is still up to either empirical observation or free and voluntary decision.

(b) Proposal 2: Arguments using generalised principles

We already noticed the mysterious attraction the "derivations by principles" had for Einstein. In fact, this mode of thinking has a strong tradition with many of the most prominent scientists in history. From elementary logic we know that there are no physical principles, which can be derived from reason alone. And, even more obviously from a logical point of view, there is no derivation of an informative, empirically law of nature from a universally generalised principle like “there is no perpetuum mobile”.

But, the history of science suggests, there is something close to it.

This something is commonly called “induction”. We can generalize from phenomenological laws. And we can continue to generalize to universal laws and to universal principles. These generalisations are as true as any other inductive reasoning may be true or not. It is a historical fact that scientists were very careful to use inductions beyond the original field of application. It took science centuries to advance from the knowledge: “there is no perpetuum mobile using mechanical elements only” to the generalized principle “there is no perpetuum mobile using any field of natural science, mechanics, electromagnetism or else.”

However, with hindsight, it proved very successful to hypothetically assume the law “there is no perpetuum mobile” in areas, where it had formerly not been applied – like Carnot used the analogy of law “there is no perpetuum mobile in mechanics” in thermodynamics. At the time, Carnot did this, nobody trusted him, because this argument is really far fetched. But, as said, sometimes – with the perception-filter of our historic awareness one could even say: often – often this far fetched inductive reasoning seems to be very successful.

In a way, proposal 2 comes down to the same message of proposal 1: it is a wise strategy in science to be bold and daring, and to advance your spirit and your known theories to fields of application, where they have never been before.

*