Reprint of:

Ulrich Kühne: “Thought Experiments and the Inference to a Coherent Explanation“, *Volume of Abstracts. 10th International Congress of Logic, Methodology and Philosophy of Science*, Florence, Italy 1995, page 238.

and:

Ulrich Kühne: “Thought Experiments and the Inference to a Coherent Explanation“, talk presented at the 10th International Congress of Logic, Methodology and Philosophy of Science, Florence, Italy, August 20th 1995, 5.20 p.m., previously unpublished.

Abstract:

This talk proposes a solution to the “paradox of thought experiments”: how seemingly new empirical insights are derived by “thought experiments” which are entirely conducted inside one’s head, so to speak, without any new empirical import. Rather than the discovery of new facts the purpose of a successful thought experiment is to transform vague intuitions about nature into explanatory standards, which may be uses later on as normative principles or formal constraints in the formation of new theories.

The author later wrote a *Dr.-phil.*-thesis on this subject: *“Die Methode des Gedankenexperiments. Untersuchung zur Rationalität naturwissenschaftlicher Theoriereformen.”* (Bremen University, submitted October 5th 2001, *summa cum laude*) an abridged version of which has been published in the book: Ulrich Kühne, *Die Methode des Gedankenexperiments.* Frankfurt am Main: Suhrkamp Verlag 2005. 410 pages, ISBN: 978-3-518-29342-3.

Author:

Ulrich J. Kuehne, KUHNE@UNI-BREMEN.DE

Section 6: Methodology

THOUGHT EXPERIMENTS AND THE INFERENCE TO A COHERENT EXPLANATION

U. J. Kühne

By common understanding thought experiments are a tool to gain new knowledge about nature by means of armchair philosophy only. The obvious consequence of this position is either a straightforward platonistic interpretation of physical laws (as, e.g., J R BROWN proposed) or the need to banish thought experimentation from the empirical sciences completely (which goes back to DUHEM). Recent statements try to evade this alternative by identifying thought experiments with less provoking scientific enterprises, like doing computer simulations (HUMPHREYS) or argumentation (NORTON). Thereby, however, they fail so far to do justice to the history of science, where the thought experiment appears to be a prominent scientific method of its own right.

Going back to the (in modern bibliography unknown) roots of the concept 'thought experiment' in the philosophical writings of the Danish scientist Hans Christian OERSTED (1777-1851), we find the foundation to a more adequate understanding of thought experiments: They are in the first place not concerned with making predictions or constructing a mental analogue of physical measurements, but with the search for a coherent *explanation* of nature. OERSTED's main philosophical interest in theorising about thought experiments was to rebuff the method of 'speculation', which was celebrated by the community of the German *Naturphilosophen* (FICHTE, SCHELLING, etc.), for getting it wrong with the empirical data, while he maintained deeply romantic ideas about the harmony between human understanding and nature. He was - in sympathy with KANT- looking for a viable middle course between the blind progress of laboratory science and the fruitless efforts of speculative metaphysics.

A rough sketch of a modern understanding of thought experiments within the OERSTED - tradition would read as follows: In trying to decide about a hypothetical or unsettled experience, thought experiments make use of intuitions about law-like relations, which were beforehand used to explain familiar experiences only. To do so the thought experiment has to present the unsettled experience in a familiar setting - thus seducing the readers to trust their

intuitions to be applicable to this setting and to decide about it. If successful, the readers are more convinced by their intuitive decision, than by the derivation from any existing theory about the given setting. There is no mystical reason, why this decision could not be proven false by real experiments. But we can draw the conclusion that, *if* the outcome of the thought experiment does *not* conflict with the empirical data, those who accept the soundness of the thought experiment obviously can explain the derived facts in coherence with their familiar knowledge and by use of principles and laws, which to them have an intuitive credibility prior to the empirical data in question. Thus thought experiments in the first place install - guided by the general goal of coherence - those principles into the scientific domain, which - were they already established in science - would reduce the thought experiment to an ordinary derivation from a given theory. Rather than telling us something about the reality outside, successful thought experiments let us extract information about *intentional constraints* to the construction of physical theories, if these theories shall qualify for explanations.

In short: while at first sight thought experiments pretend to succeed with the derivation of nontrivial facts out of reason alone, their rôle in the progress of science is to transform common intuitions into intentional constraints on well-formed theories - as can be demonstrated by the history of, e.g., the superposition principle, the law of inertia, or the demand of LORENTZ - invariance and their respective thought experiments

———

Thought Experiments and the Inference to a Coherent Explanation

Florence, 20/08/95, 5.20 p.m.

By common understanding thought experiments are a tool to gain new knowledge about nature by means of armchair philosophy only. However, our common understanding of the facts of nature states that these are contingent, and that neither conceivability nor logic, mathematics, analyticity or any combination of these suffice to decide about the truth of an empirical matter of fact. The decision about a matter of fact can, by the standard doctrine of empiricism, only be done by looking at nature with our senses, making observations and measurements.

Thus the so-called "paradox of thought experiments" arises, which can be phrased by the question: Where does the information come from which great scientists in history regularly claim to have derived by thought experiments (and which too often turned out true to be dismissed as a chance guess)?

It has, by a committed minority of authors, been argued that thought experimentation is in fact the long-looked-for instrument (or rather: sense-organ) for the exploration of the a-priori order of nature. By way of doing thought experiments, they claim, we could forward directly to the truth without dwelling on the tedious business of induction from the dirty data of unreliable real experiments. - This interpretation of thought experiments has, however, until now not been developed in enough detail and with rigour enough to convince a sceptic – and the lack of any demarcation criteria between sound thought experiments in accordance to this claim and sheer nonsensical speculation impedes me from any further elaboration.

But the opposite interpretation, that is to say: thought experiments are just a fake, deceiving their audience by disguising their lack of substance with fancy suggestive stories, I regard unsatisfactory, too. This is because it cannot answer to the rôle thought experiments always played in the development of scientific theories. I'll come back to this point in a minute.

Even though it is now at large accepted that thought experimentation is a common and legitimate method within the empirical sciences, (or I should

rather say: that the practical scientist wouldn't stop thought experimentation if philosophers decide they were illegal) the problem to make sense out of them within a plausible philosophy of science has not yet been resolved. The usual way to evade the alternative of the just-mentioned extreme positions is to identify thought experiments with less provocative scientific methods. For instance, if Norton proposes that thought experiments are just arguments, the claim that new knowledge about nature can be gained by them without leaving the armchair is not surprising at all. However, it is then hard to understand, why thought experiments always "invoke particulars irrelevant to the generality of the conclusion" - Why do thought experiments always come in the full ornament of detailed stories about counterfactual situations, if the core-argument suffices?

Mach, to give another example, explained the predictive power of thought experiments with reference to hidden stores of empirical knowledge. He writes that for instance the grammar of our language has, by an evolutionary process, been adopted to the physical reality. Thus, when thought experiments look like an a-priori reasoning, they in fact restore, according to Mach, the empirical data, which went into the formation of human modes of thinking. But here again we should ask: Why shouldn't it be enough to restate the hidden knowledge explicitly, and why bother with thought experiments? - In general: if one identifies thought experiments with something well known, the whole concept of "thought experiment" becomes superfluous. It is a matter of conceptual hygiene to banish the word "thought experiment" from the methodological terminology, if it can be replaced in all instances by more familiar names - like "argumentation" or, in other cases, by "simulation", "illustration", "idealisation" or "model based reasoning" and so on.

Time allows at this conference only to state a problem (which I just did) and answer it. I want to apologise for the unphilosophical shortness of my answer.

The thesis I want to put forward is that thought experiments are in fact a necessary part of scientific methodology, but not concerned with the derivation of new knowledge about nature by means of armchair philosophy, but with the explanation of nature (which is a far more natural thing to do in an armchair). The impression that often in the history of science thought experiments were capable to predict the outcome of real measurements (which, at the time of the thought experiment, were for technical reasons impossible) results - according to this thesis - from the nontrivial interdependencies of

prediction and explanation in scientific methodology, which needs a detailed analysis. But before I add a few words to this problem, let me name some reasons, why I think thought experiments deal with the development of explanations of nature (as opposed to: predictions of nature):

One reason is (of course) that the above mentioned "paradox of thought experiments" is otherwise unsolvable: If thought experiments are not just arguments on the basis of old knowledge and can predict contingent matters of fact, then it would be plain necessary that platonism is true or that we are cheated. And it would be most probable true, I think, that we could never decide.

An other (perhaps weak) reason is that the concept of thought experiment was actually invented as the name of the method to reach to "real" explanations of nature (as opposed to mere phenomenological laws to describe nature) This was done by the Danish chemist, physicist and philosopher Hans Christian Oersted (living from 1777 till 1851). However Oersted's philosophical writings have been almost totally ignored in this century and for lack of time I have to do so, too. But it should be remarked that at the time of around 1810, when Oersted developed his scientific methodology based on four principal scientific tools (namely: 1: getting everyday experience, 2: making observation, 3: planing and realising experiments, and last, but not least, to construct thought experiments) there already existed the philosophical name for an scientific method to predict nature. This was the so-called "speculation", and "speculation" was taken perfectly seriously by a large part of practical scientists at this time as a method to replace real experiments at some stage of philosophical progress. Oersted was among the first to recognise that this claim would discredit the whole of philosophy of science, if it is not reduced to a reasonable level. But Oersted was not heard at and the so-called "Romantic Naturphilosophie" did discredit the whole of philosophy of science for a long time, till the advance of empiricism at the beginning of this century. But the dyed-in-the-wool empiricism cannot deal with the apparent rationalistic element of thought experimentation in the methodology of practical scientists. Thus we are today on the other side of the bridge at pretty much the same situation of Oersted, when he wrote about thought experiments. - So much for a rush through the history of methodology in three sentences.

As a third reason why I see the objective of thought experiments in the goal to explain nature I want to state how I see their role in the history of science.

If one does only look at the truly spectacular thought experiments and discards all boring ones as being just terminological inflation, one can find the following pattern:

Thought experiments become only relevant during the Kuhnian "revolutions" of science. More than this: They seem to constitute an indispensable part in the process of discrediting an old and developing a new theory. This holds true, even if the empirical data at hand completely justifies the change of theory. For an illustration of this claim we could have a look at the advance of quantum theory: At first sight it looks like a counterexample, since the formulation of the mathematical apparatus of the theory has been done solely to "save the phenomena". While there are plenty of thought experiments during the development of the mathematics of, for instance, the theories of relativity, the history of quantum theory till around 1925 can certainly be exhausted without the reference to a single thought experiment. But, as if every scientist agreed that this constitutes a deficit in the quality of quantum mechanics, the thought experiments were handed in later, e.g. Heisenberg's gamma-ray-microscope, Schrödinger's cat and so on. --- To my interpretation the indispensability of thought experiments in the development of theories simply reflects the fact that scientific theories serve two completely separate goals. One is to predict empirical matters of fact. For this purpose it probably suffices to have a mathematical formula, which was derived from real experimental data by some method of induction. But the other goal of theories is to explain the phenomena. To serve this purpose, the theory needs corrobation by thought experiments independent from the results of real experiments.

As a second result of an analysis of the great paradigmatic thought experiments in the history of science one certainly finds full agreement with Norton that they can be reconstructed as mere arguments. However, one notes that in all of the paradigmatic thought experiments at least one of the premises, the reconstructed argument rests on, was at the time, the thought experiment was introduced for the first time, not explicitly accepted by the scientific community. They even sometimes stood in strong opposition to what was commonly accepted in science at the time, the thought experiment was thrown into the discussion. And one notes that these premises are no tautologies. That is: it is at least conceivable that these premises were false. But history shows that these premises in the time after the acceptance of the thought experiment constitute the more or less unreflected conceptual basis of the new theories. Their standing inside the new theoretical framework is of

such a strength that the platonists like Koyre or Brown apparently are deceived to regard them a-priori necessities. But it is more appropriate to use Toulmin's phrase "ideals of natural order" to characterise the premises, which by the means of thought experiments are introduced into the conceptual foundation of scientific theories.

What does this have to do with explanations? Obviously very little, if one insists - in the tradition of Hempel - on the structural identity of prediction and explanation. However based on van Fraassen's pragmatic interpretation of explanation it sounds, to my opinion, perfectly natural to argue in the following line: Something is an explanandum, if one feels like asking a why-question about it. One feels like asking a why-question, if the proposition sticks out of what is regarded as the normal course of nature. And the explanation should give the reason, why - as opposed to the impression at the first sight before asking the why-question - the matter of fact in the explanandum actually can be understood to be in perfect harmony with what the person asking accepts to be the uncontroversial normality. In short: Scientific explanations in general deal with the coherence of knowledge. And in special they have to prove the coherence of scientific theories with the familiar knowledge of our everyday experience.

Now, how does a thought experiment work? A thought experiment confronts us with a detailed story about some hypothetical setting. For being worded in such familiar terms one is convinced that our familiar knowledge is applicable to this setting. We apply the same modes of thinking to this setting which we normally use to explain an element of our everyday world without reference to a mathematical theory. If, for instance, I would be asked to explain, why a heavy stone falls faster than a light one (what, in fact, isn't true), I would argue that this would be (if it were true) in coherence to our knowledge, namely that there pulls a stronger force of gravity on a heavy stone then on a light one (as we feel, when we have to carry them) and that stronger force makes a body faster than a weak one (as we know from the fact that a chariot with two horses can be faster than a chariot with only one horse pulling) - The role of thought experiments in the development of science is to test, whether a mathematical theory at hand is coherent with our natural reasoning. And this is to say, I claim, to test whether a mathematical theory has an explanatory power. For instance, Aristotle's theory of free fall is in line with the just sketched argument. He claims that the velocity of freely falling bodies is proportional to their weight. But then Galileo and Benedetti noticed an incoherence of Aristotle with an other element of our familiar expecta-

tions: While we do expect a chariot with two horses to be faster than one with only one horse, we do expect two chariots with one horse each not to become any faster, if we tie both chariots together to form a single chariot with two horses. And our familiar modes of thinking demands that this familiar knowledge applies to a situation of two freely falling stones, if we tie them together by a bit of string to constitute a single big stone. Thus, if Aristotle's law of free fall would be correct (what it isn't, but what could in a possible world be the case) two stones of the same size falling in close distance with the same speed would suddenly accelerate, if they are connected by a little bit of string. And this fact in the possible world of Aristotle would be incoherent to our natural expectations.

But it is of course by no means granted that a theory with a higher explanatory power than an other is necessarily true. Neither that our intuitive expectations will not lead us astray. The fact that some thought experiments have been able to predict the outcome of a real experiment simply reflects that it is a good and time-honoured strategy to suppose a general coherence in nature when one searches for theories with predictive power.

* * *