THOUGHT-EXPERIMENTS AND MODAL LOGICS

Dr. Wulf REHDER

Introduction

Human reason, writes Immanuel Kant in his preface to the first edition of Critique of Pure Reason, suffers from a typical dilemma: it bothers itself with questions which reason cannot reject; because reason itself, by its very nature, poses those questions; but which reason, on the other hand, cannot answer. Eventually, reason transcends all empirical facts and does not any longer acknowledge experimental standards.

Physicists, however, hold that reason can answer, if not its own intrinsic problems, at least certain questions about the «outer nature» – or rather: physicists turn the table by putting the question to Nature via test-experiment. Thus it seems, as if the monologue of reason has been replaced by a dialogue with Nature. And yet, it has been a long way from Bacon's «interpretatio naturae» to science, where «asking» or «interpreting» Nature is dominated by exploiting or at least taming and controlling Nature. We must not forget, however, that at the same time an unprecedented theoretical effort has widened and deepened our understanding of Nature.

This paper intends to go back even one more step into the realm where those trying questions are formulated for the first time, to go back to sources of possible theoretical knowledge not concerned too much about touchstones from experience – back to thought-experiments. After giving some typical examples from Aristotle to Einstein, we sketch some opinions ventured by Mach, Meinong, Popper, and Duhem about general aspects of thought-experiments. In our last and main section, we try to give an informal description and a formal characterization in the context of Kripke's semantics of possible worlds, linking thought-experiments and modalities.

Some typical thought-experiments

One of the first and at the same time most influential thought-experiments can be found in Aristotle's De Coelo 296^b 22. It was meant to refute the hypothesis of a moving earth by a reductio-argument which was reformulated by Galilei in his persuasive tower-argument: if the earth moved, a rock falling from the top of a tower would be left behind and it would hit the ground only at a distance from the foot of the tower. (1)

It was again Galilei who, after presenting several similar examples, seemingly corroborating Aristotle's point, introduced a new mathematical conception based on Kepler's ideas and on his own anti-common-sense principles of inertia, relativity, and superposition of movements. His popular proof of constant speed for all freely falling bodies, however, is ingenious common-sense: assuming Aristotle's «law» that heavy bodies have higher velocity than lighter ones, leads to an absurdity: just tie a heavy rock and a small pebble together. As the joint package becomes heavier even than the big rock, it ought to pick up speed. On the other hand, the smaller velocity of the lighter pebble decelerates the downward movement (principle of independent superposition). Like Aristotle before him, Galilei did not bother to actually carry out his thought-experiment.

Another famous and much admired thought-experiment was Stevin's derivation of the law of the inclined plane: he imagined a triangular block over which he laid a chain with fourteen balls of equal size, weight, and distance from each other. Since such a homogeneous chain obviously does not move around by itself, Stevin was able to infer his basic laws from this state of equilibrium. (2)

We take a big step from Stevin's and Galilei's ages to modern times where Einstein criticized Galilei's idea of inertial systems and New-

(1) G. GALILEI, «Dialogi», engl. trans. Berkeley (1953), p. 126. For the tower-argument and also the following discussion of Galilei's method cf. P. Feyerabend, «Against Method», chapters 6 and 7, London, NLB (1975). There the relevant literature is quoted extensively. A shorter and more balanced account of Galilei's rôle in modern science may be found in E.J. Dijksterhuis, «Mechanization of the World Picture», Galaxy Books, (1970), part IV, sections 97-123.

(2) Stevin's basic idea is admirably described in Dijksterhuis's book, cf. note 1), part IV, sections 60-69.

ton's theory of gravitation through his simple and brilliant thought-experiments or, as Einstein used to say, «idealized experiments».(3) There is, for instance, his well-known definition of relative synchronicity of time by using light-rays and trains moving on straight tracks. This thought-experiment served as a refutation of Galilei's principle of relativity. It used the speed of light as a constant maximum speed, and it vindicated at the same time the Lorentz-transformations as the true rules for superposition. Later, for his theory of general relativity, Einstein chose rather witty situations with observers experimenting in accelerated elevators watched from the outside by their curious colleagues. In order to explain certain features of curved spaces, he used «geometrical experiments», describing two-dimensional «humans» trying to do physics on a large curved plane. These idealized experiments are not designed to be performed or to be found anywhere; they serve as illustrations, as a heuristic introduction of new concepts, or as a critical test of old and new theoretical approaches.

Let us now make a few remarks on three tentative «theories» of thought-experiments which have been proposed by the physicist and historian Ernst Mach, the philosopher Alexius Meinong, and Sir Karl Popper.

Mach(4) seems to have been the first to coin the term «thought-experiment», or, in German: «Gedankenexperiment», a word which even appears in Merriam Webster's Unabridged Dictionary. In principle, thought-experiments obey the same basic pattern as empirical experiments: the method of systematical continuous variation of initial conditions and other relevant circumstances. Parallel to these variations, our expectations of the outcome of the experiment vary, so that there is a certain continuity relation («adaptation») between the (expected) facts and our thoughts: «A rock drops to the ground. Let

(3) Einstein's «idealized» experiments can be found in every modern book on Mechanics. A non-technical account was given by Einstein, L. Infeld, «Die Evolution der Physik», Rowohlt rde 12, Hamburg (1956), see esp. pp. 105-164.

(4) E. MACH, «Erkenntnis und Irrtum», Leipzig, 2nd ed. (1906), pp. 183-200 «Uber Gedankenexperimente». There is also a recent English translation, ed. by B. McGuinness, «Knowledge and Error», transl. by P. Foules, Vienna Circle Coll. 3, Reidel (1975). I have not seen this translation yet, so all quotations are my own translations from the German edition of 1906.

its distance from the ground increase. Now, it would do violence to our commonsense to counter this continuous growth with a discontinuity of our expectation. Even at the distance of the moon the rock does not loose its tendency to fall. A big rock falls as a small one does. Assume the rock to grow as big as the moon. The moon, too, tends to fall towards the earth. Let the moon grow to the size of the earth. Now it would be inconsistent to assume that only one attracts the other, but not vice versa. Hence attraction is mutual.» Again, there is no chance of really performing the described experiment, and Mach therefore calls it only an abstraction or idealization.

Alexius Meinong([5]) offers «an utterly destructive criticism of Mach's 'Gedankenexperiment'», as Bertrand Russell wrote in Mind.([6]) In particular, Meinong argues against experimenting *with* thoughts, something that in his opinion is apt to lead to illusions and should be studied by experimental psychology, which in those days was going very strong in Austria. At the most, a thought-experiment poses – metaphorically speaking – a question to Nature, but a definite answer can only be given by an empirical experiment. So in the end, Meinong merely concedes a non-committal thinking *about* an experiment as the proper rôle of a thought-experiment.

Sir Karl Popper,([7]) in his classic Logic of Discovery, is extremely sceptical about the so-called «defensive» or «apologetic» usage of thought-experiments such as Bohr's defense of the uncertainty principle against Einstein-Rosen-Podolski's well known paradox. (At least this is Popper's reading of the controversy.) Popper only approves of the critical, falsificational aspect of certain thought-experiments, of which Galilei's reductio-argument involving two connected rocks as sketched above, is the best paradigm. Popper completely neglects any

([5]) A. MEINONG, «Das Gedankenexperiment», § 15 of his work «Uber die Stellung der Gegenstandstheorie im System der Wissenschaften», repr. pp. 273-283 of vol. V of Alexius Meinong Gesamtausgabe, Hrsg. R. Haller, R. Kindinger, gem. mit R.M. Chisholm, Graz, Austria (1973).

([6]) B. RUSSELL, Mind n.s. 16 (October 1906), pp. 436-9. Russell's review of Meinong is reprinted on pp. 89-93 in «Essays in analysis», ed. by D. LACKEY, London, Allen and Unwin (1973). There are also, on pp. 17-76, two more critiques of Meinong's so-called «Gegenstandstheorie».

([7]) I have used the 3rd German edition of «Logik der Forschung», J.C.B. MOHR, Tübingen (1969), chapter XI, pp. 397-411.

heuristic relevance, and looking closely at some typical thought-experiments (e.g. Einstein's) shows clearly that physicists use defensive arguments quite freely. Einstein, for instance, staunchly defended his principle of constant speed of light against Miller's allegedly successful proof of the ether.

Last but not least, Pierre Duhem([8]) could also be mentioned as an adversary of «fictitious» experiments leading to absurd quantities like G. Robin's corps témoins and, in the whole, to a contamination of his clear-cut separation of symbolic representation of physical phenomena through mathematical theory on one side and «crowning» experiments on the other.

After this brief survey of mainly critical appreciations of certain aspects of thought-experiments, we intend to start at a new beginning in our next, more systematic, section. Taking into account the above tentative remarks by Mach, Meinong, and Popper, and using the examples from the beginning of this section for illustrations, let us try a formal analysis of thought-experiments by means of modal logics and possible worlds.

Thought-experiments and possible worlds

In order to discuss thought-experiments in the context of modalities, we shall make use of a very suggestive game-theoretical interpretation of Kripke's possible-world semantics.([9]) As Kripke's description may be understood as a precise reformulation of Leibniz's pays des possibles, some of the ideas from the Elementa juris naturalis and Generales Inquisitiones will serve to illustrate our point.([10])

To begin with, let us introduce a very general scheme which will be

([8]) P. DUHEM, «The Aim and Structure of Physical Theory», Atheneum, N.Y. (1962), pp. 201-205.

([9]) S.A. KRIPKE, «Semantical analysis of modal logics I, normal propositional calculi», Z. f. mathem. Logik u. Grundlagen d. Mathematik, VEB Verlag, Berlin (1963). I have used G.E. Hughes and M.J. Cresswell, «An Introduction to Modal Logic», Methuen, London (1968), see esp. chapter 4, and appendix 5 at the end of the book.

([10]) For Leibniz's modal logics cf. H. POSER, «Zur Theorie der Modalbegriffe bei G.W. Leibniz», Stud. Leibnitiana Supplementa, Band VI; F. Steiner, Wiesbaden (1969), esp. pp. 16-42, 61-75.

specialized later. The letter W denotes a set whose elements w_1, w_2,..., the so-called possible worlds, may themselves consist of elements or entities. In what follows, w_1 will always be our real world in which we live and make thought-experiments. Amongst the rest of W are included, for instance, a «world without friction» and a «Platonic world of ideal objects», in particular «Hilbert's complex vector-spaces of quantum-theory», and also Einstein's «geometrical fantasy-world of two-dimensional creatures», and, of course, his «4-dimensional space-time-world of general relativity», and many more. Worlds of such a wide variety would hardly show very many common features, but by giving a certain structure to W, we ought to be able to relate several possible worlds.

To achieve this, let there be defined a binary relation R in W, and if two worlds w_i and w_j are related, i.e. w_iRw_j, we say that «w_j is accessible from w_i with respect to R». R is always assumed to be reflexive, meaning that every world is accessible from itself. Accessibility is by no means self-explanatory, and we use the word here only to appeal to the reader's intuition: he is invited to read w_1Rw_2, say, as « a physicist in our actual world w_1 is capable of imagining a world w_2 without friction», and in this sense w_2 is accessible to him.

Propositions p about w_2 are either true or false according to a truth-valuation in w_2. p = «falling bodies in a vacuum and in a homogeneous field of gravitation g obey Galilei's law $s = 1/2gt^2$» is to be valuated as «true» in the world w_2 of classical mechanics, whereas Aristotle's «the velocity of falling bodies is directly proportional to their weight» has the value «false» in w_2, even though his «law» may sometimes, e.g. for light objects like leaves as compared to rocks, be approximately verified in real life, i.e. in w_1. Strictly speaking, a valuation would have to be defined recursively in order to comply with some logical requirements, with which we shall not concern ourselves, however. Let us proceed to obtain a description of thought-experiments in the above set-up.

We assume that everything that can be thought of and asserted in a thought-experiment, can also be said and written down, so that we may condense the essence of a thought-experiment ε in a proposition p (ε). This proposition puts a physicist's question to Nature «do all bodies fall at the same speed?» already in its corresponding assertion form «all bodies fall at the same speed». We note that the above

question and its propositional form have the same Fregean sense («Sinn»): the thought («Gedanke») expressed in them is identical. Thus, the gist of a thought-experiment ε can be adequately given through $p(\varepsilon)$.

Definition: ε is a (W, R, V) thought-experiment in our reference world w_1, if there is at least one possible world w_i in W such that w_i is accessible from w_1:
$w_1 R w_i$, and that $p(\varepsilon)$ is true in w_i, abbreviated by the valuation $V(p(\varepsilon), w_i) = 1$.

Looked at in this way, a thought-experiment is defined in terms which call to mind Leibniz's phrase from his Elementa juris naturalis: possibile est quicquid quodam casu (scil. verum est), so that in this sense a thought-experiment is a «possible» experiment.

The next step in our discussion will lead us to an informal exposition of three types of thought-experiments which, in a subsequent more formal approach, can be shown to represent the three modal systems S5, S4, and B, if we define the respective accessibility relations appropriately.

(1) A thought-experiment as «thinking *about* an experiment» refers to an imaginable experiment, an experiment that is conceived to be carried out. To use Leibniz's words, it is an experiment which «potest fieri», something that is capable of being actualized. Galilei is, I think, a main exponent of this type of thought-experiment - regardless of the fact that very often he himself did not care to actually perform them afterwards.

(2) To do an experiment «*in* one's mind or thoughts» applies to «idealized experiments» which are abstractions from experience. The corresponding proposition must not contradict *vérités de fait* nor the logic underlying physical theories. Not all contingent conditions need to be taken into account, however: technical difficulties, human infirmities, the imperfection of the apparatus or else unwanted interruptions or hindrances like friction may be neglected. Thought-experiments of this category are not primarily conceived in order to be carried out eventually; they are meant to give a good, albeit idealized picture of the relevant traits of the physical problem under investigation, and thus it serves rather as an abstract test of consistency for (parts of) a theory.

As an example, we include Einstein's train-thought-experiments here which were not designed to be performed in reality, but were envisaged as a criticism of Galilean inertial systems. At the same time, these «idealized experiments» functioned as a vindication of the Lorentz-transformations and as a preliminary test for the principle of a constant velocity of light. A plethora of further examples is provided in Mach's section on thought-experiments in Knowledge and Error, and another famous more recent instance is Heisenberg's super-microscope more powerful than any electronic microscope. Heisenberg's microscope cannot be built, but it is not logically impossible.

(3) Experimenting «*with* thoughts» - here the thoughts themselves are the stuff with which we do the experiment. This sort of thought-experiment is neither intended as a mental blue-print for an actual performance of an experiment in a laboratory, nor is it merely an idealized abstraction from reality «forgetting» certain contingencies. No, here thinking itself – by way of an adaptation of thoughts (Mach) – experiences hitherto unknown things: *mens potest experiri,* or even: *cogito, ergo possum experiri,* versus the *experimentum quod potest fieri* of type (1). Of course, we do not mean sense-experiences, but logical or mathematical intuition and understanding made possible by trained intellectual skills and mental creativity. A mathematician or theoretical physicist may experience structural insights by «playing» with symbols which make sense to him without always having a «meaning» (à la Frege) in the real world. He experiences, we might say, with what Wittgenstein called logical or mathematical manifolds (Tractatus 4.04). Thus, experimenting *with* thoughts in this sense admits Einstein's «geometrical experiments» employing two-dimensional creatures and their physics, and other mathematical modeling.

In the following, the above informal characterization of three types of thought-experiments will provide the background for a more formal treatment along the lines of Kripke's semantics of possible worlds such that the categories (1), (2), and (3) represent the modal systems S5, S4, and B, respectively.

(1)-S5: The constitutive element of type (1) thought-experiments is their practicability: *potest fieri*, tomorrow, or after the rain stops, or in Canada. Accordingly, a binary relation R_1 in W is adequate if it relates w_1 to worlds w_i which are physically equivalent to w_1: in w_i, e.g. tomorrow, the same physical laws hold, and the relevant equip-

ment for measurement (clocks, tape-measure, etc.) are essentially the same (centimeters and inches may be used equivalently). A suitable valuation V_1 has to be introduced over all equivalent w_i where our experiment can be performed, and for ε to be called a (W, R_1, V_1) thought-experiment, there must exist an accessible world w_j in which $p(\varepsilon)$ becomes true. R_1 is an equivalence-relation defining a partition of W into a set of worlds w_i equivalent to w_1, and the rest. Hence (W, R_1, V_1) is a model for S5.

(2)-S4: Whereas in (1) physically equivalent worlds are related by – as we might say – an «isomorphism» R_1, we define a weaker accessibility relation R_2 for type (2) thought-experiments as follows: $w_i R_2 w_j$ if w_j is a homomorphic image of the (somehow structured) world w_i: $w_j = h_{ij}(w_i)$. The structural homomorphism h_{ij} «forgets» certain inconvenient interferences like friction, technical shortcomings, and human imperfection and mistakes. On the other hand, h_{ij} preserves all physical laws relevant for the experiment in question without translating all of w_i-physics into w_j. ([11]) A suitable valuation $V_2(P(\varepsilon), w_i) = 1$ for Galilei's example ε of freely falling bodies may be read as: «if w_i possesses a homogeneous field of gravitation g and no friction, the proposition 'a freely falling mass-point has velocity $v = gt$ after the elapse of time t' is true.» Since the composition of homomorphisms is again a homomorphism, R_2 is transitive, but generally not symmetrical, which means that not all properties from an idealized world w_i can be found in our real world w_1! (W, R_2, V_2) is thus a model for S4.

(3)-B: Thought-experiments of this third category bear only a faint resemblance to practicable experiments. Not only do we leave the realm of tangible things when we talk about «experimenting *with* thoughts» – even physical laws may be suspended.

Thought-experiments envisaged in type (3) aim at tentative mathematizations, e.g. at Einstein's «geometrical experiments». Observable facts are of secondary importance and yield to symbolic representations by formal systems which do not merely reflect a contingent real world through an idealized purified image, but which are new worlds themselves. Furthermore, an extensional characteri-

([11]) We remark that this homomorphism h_{ij} is, by E. NOETHER'S first theorem on isomorphisms, canonical, i.e. essentially uniquely determined by those circumstances in w_i to be «forgotten» in w_j.

zation of possible experiments by means of sets of entities and maps between them (as in (1) and (2)) is problematic, for the very reason that here «sense» and «meaning», «insight» and «understanding» play a dominant role. Therefore, I think, an intensional formalization would be more adequate, which admits talking of «meaning of names» and of different «senses of propositions». A matching accessibility relation has to be defined in semantical terms. This, indeed, can be done by employing Montague's theory of reference: senses are functions from possible worlds into sets of entities E, or truth-values $\{1,0\}$, or other types of designates.([12]) It is enough to restrict ourselves to Montague's types $<s,e>$ (approximately Frege's «Bedeutung» or «meaning») and $<s,t>$ (Frege's «Sinn» of propositions, or the «Gedanke» (thought) expressed in them).

In this intensional framework a semantical accessibility relation $R_3 = R_3(S)$ in W can be introduced, depending on types from $S \subset \{e,t,\ldots\}$: $w_i R_3 w_j$ if there exists a $\tau \in S$ such that for all functions f from W into the corresponding set of designates of type τ the «senses» are identical: $f(w_i) = f(w_j)$. For the basic category $<s,t>$ accessibility of w_j from w_i indicates that a proposition p being true in w_i holds true in w_j also; or, for $<s,e>$, a (different) accessibility ensures that names of things *mean* (stand for, «bedeuten») the same entity in w_i and w_j.

This latter meaning-identity is important for an astronomer, say, who wants to make thought-experiments (theoretical inferences from his theory) for the planet Venus: he might base his calculations, assumptions, etc. on different aspects, viewing Venus as an ideal mass-point, or as a line in space-time, or as a certain field of gravitation, or as the morning-star, or as the evening-star. Yet all these different senses have to coincide in one and the same «meaning», i.e. designate the one and only star Venus. R_3 is, for any $S \subset \{e,t,\ldots\}$, reflexive and symmetrical, but in general not transitive, if S contains more than one element. Consequently, (3) provides a model for Brouwer's system B.

The thesis T34: $MLp \supset p$ which distinguishes Brouwer's system

([12]) R. MONTAGUE, «Universal Grammar», in «Formal Philosophy. Selected Papers of Richard Montague», ed. with an introduction by R.H. Thomason, Yale University Press, New Haven and London (1974, 2nd printing 1976), pp. 222-246.

$B = T + T34$, admits a material implication from a possibility (MLp) to an actuality (p). Let us formulate the crucial T34 by explicitly using the definition of our intensional accessibility relation R_3 as given above: there is a possible world w_i which coincides with our real world w_1 in a certain sense τ, and Lp is true in w_i. From this we infer by using the definition of the necessity operator L that p is true in all those worlds w_j which are accessible in some sense from w_i. Our actual world w_1 is among them (since R_3 is symmetrical), and so p is a fact in w_1.

In this way, the modal system B permits a valid inference from possible «fantasy»-worlds back to facts in the real world. This is of great importance for the conclusiveness and applicability of thought-experiments. It means, e.g., that «logical experiences» or «experimenting with thoughts and symbols» may either be transferred via R^3 into reality, thereby preserving the sense, or else via material implication and modus ponens.

Technische Universität Berlin] Dr. Wulf REHDER