**Pre-press copy: BUZZONI, Marco: Kantian Accounts of Thought Experiments. In J.R. Brown, Y. Fehige, M. Stuart (eds.). The Routledge Companion to Thought Experiments, New York, Routledge, pp. 327-341 (in press)**

**18**

Kantian Accounts of Thought Experiments

*Marco Buzzoni*

**1 Introduction**

Kant did not distinguish between thought experiment (TE) and real-world experiment (RE), nor did he use the German term *Gedankenexperiment.* But he is of great importance in the history and theory of TEs. Many aspects of Kant's critical philosophy are connected with the concept of TE, including the definition of scientific experiment, the distinction between knowing and thinking, and the fact that "experiments" of pure and practical reason embody the specific method of philosophy (Section 2).

However, the main reason Kant plays such an important role in the history of the concept of TE is because the expression was first used not by Mach, as was believed until a few decades ago, but the Danish scientist Hans-Christian Ørsted, who introduced it (not only in Danish: *Tankeexperiment,* but also in German: *Gedankenexperiment*; Ørsted **1811** [1920]) in order to clarify the relation between mathematics and physical knowledge in Kant. Section 3 will be devoted to Ørsted's account, the first explicitly Kantian theory of TEs.

As we shall see, Ørsted wavers between two positions: sometimes he claims that TEs are merely the hypothetical aspect of REs while other times he assigns to them the capacity of determining a priori some fundamental laws of natural science. The latter position recalls the "material" aspect of the Kantian a priori, a notion which was abandoned by the philosophy of science at the end of the nineteenth century.

This is probably the main reason Ørsted's point of view had no real influence on the historical development of the concept of TE and why some recent Kantian or neo-Kantian accounts of TE depart from Ørsted's concept of the a priori. I, on the other hand, have defended an interpretation of the a priori as purely functional by retaining the necessary and universal character of Kant's a priori while rejecting the material a priori, since content can be given only by experience. On this view, TE and RE are complementary in a sense similar to the complementarity of form and matter in Kant: TEs without REs are empty; REs without TEs are blind (Section 4). On the other hand, Yiftach Fehige has defended a relativized and contingent notion of the a priori which has been recently advocated by many authors, notably by Michael Friedman. Unlike myself, Fehige rejects the universality and necessity of Kant's a priori but

retains the idea that the a priori, as a constitutive element of experience, is endowed with material content that may be made explicit by thought experimentation (Section 5).

**2 Kant's concept of scientific experiment and the**

|  |  |
| --- | --- |
|  | **“experiments of pure” reason”** |

In the present section, we shall briefly discuss a few of the many aspects of Kant's critical philosophy that are intimately connected with today's theories of TE, such as the definition of scientific experiment, the distinction between knowing and thinking, and the fact that Kant held that the "experiments" of pure and practical reason embody the special method of (transcendental) philosophy.

According to Kant, scientific experiments are "questions" put to nature (KrV B xiii—xiv, AA 10, lines 17 and 27). These questions are guided, as it were, by hypothetical thinking; as Kant puts it, the experimenter is not viewed "in the character of a pupil who listens to everything that the teacher chooses to say, but of an appointed judge who compels the witnesses to answer questions which he has himself formulated" (KrV B XIII, AA III 10).

This definition is important for two reasons. First, because it assumes an internal conceptual space, a kind of space of reason where immediate intuition is questioned from a conceptual standpoint. Secondly, because it implicitly assumes that, in some sense, answers may be hypothesized before the verdict of experience. Later on, in the Preface to the *Metaphysical Foundations of Natural Science,* Kant radicalizes the autonomy of this conceptual space to the point that all genuine natural science is deemed to include a "pure part" *(reiner Theil),* which is the object of a "pure rational knowledge" of "the a priori principles of all the other explanations of nature" (cf. AA, IV, Vorrede, 469, lines 2-22). As we shall see, the relative autonomy of this space of reason paved the way for Ørsted's theory of TEs.

Many of Kant's conceptions bear a close relation to today's notion of TE. Among the most important of Kant's conceptions in this respect are the distinction between "knowing" *(Denken)* and "thinking" *(Erkennen)* and the notion of "experiments of pure reason" *(Experimente der reinen Vernunft).*

With respect to Kant's well-known unity and distinction between knowing and thinking, I refer to Kant's key objection to the ontological argument, namely the confusion between "to exist" and "to be conceived of as existing," the confusion between actual existence and existence only in thought ("gedachte[r] Existenz") *(Der einzig mögliche Beweisgrund zu einer Demonstration des Daseins Gottes,* AA II 156, line 14). Parallel with and related to this distinction is that between "to exist only in thought" *(existiren nur in Gedanken)* and "to have only in thought" *(nur in Gedanken haben).* These expressions are used by Kant in the many passages in which he speaks of reason's ability to think something as merely possible or to abstract material from formal (a priori) components (for example, in order to obtain the idea of a pure, non-empirical and absolute space; cf. MAN, AA, W, 481-2 and 506-7).

Finally, the concept of "experiments of pure reason" introduced by Kant in the Preface to the second edition of the *Critique,* has important connections both terminologically and conceptually with today's debate on TEs. This concept addresses a problem which, with very few exceptions, has received little attention in the literature on Kant in general (cf. Kalin 1972; Westphal 2003, 2004; Shi-Hyong 2008; Fulkerson-Smith 2013) and on TEs in particular (see Kühne 2005; Buzzoni 2011; Fehige 2012; Fehige and Stuart 2014).

Experiments of pure reason are intended to clarify the method of transcendental philosophy: this is why they are central to a Kantian appraisal of philosophical TEs, a task yet to be accomplished. In philosophy, unlike the natural sciences, one cannot test claims by experimenting with the relevant objects (cf. KrV B XX—XXI Anm., AA III 14 Anm.). Accordingly, for transcendental arguments not to be arbitrary there must be a criterion which enables one to distinguish true arguments from false ones. This criterion, given by the "method" of the experiments of pure reason, consists in considering objects "from two different points of view," namely "as objects of the senses and of the understanding for experience *[für die Erfahrung]"* on the one hand and "as objects which are thought merely" on the other. If, when considering things from this dual standpoint, we find that "there is agreement with the principle of pure reason," while if we adopt a single point of view there arises an "unavoidable self-conflict" of reason, then "the experiment decides in favour of the correctness of this distinction." (KrV B XVIII—XIX Anm., AA III 13 Anm.)

In other words, an experiment of pure reason consists in showing that a certain distinction (say, the distinction between things as appearances and things in themselves) is true because it is both a necessary and a sufficient condition for avoiding a contradiction of pure reason with itself:

The *analysis of the metaphysician* separated *[schied]* pure a *priori* knowledge into two very heterogeneous elements, namely, the knowledge of things as appearances, and the knowledge of things in themselves; his *dialectic* combines these two again, in *harmony* with the necessary idea of the *unconditioned* demanded by reason, and finds that this harmony can never be obtained except through the above distinction, which therefore is the true one *[welche also die wahre ist].*

(KrV B XX-XXI Anm., AA III 14 Anm.)

From this point of view, Kahn's thesis that Kant's transcendental arguments "can be characterized [...] as Gedankenexperimente" (Kalin 1972, 322) can be applied to the entire *Critique of Pure Reason.* The whole book can be seen as an experiment of pure reason, a philosophical TE whose truth is guaranteed by the fact that human understanding loses itself in antinomies when it ventures beyond the limits of possible experience and attempts to deal with things in themselves (Buzzoni 2011, 105).

Experiments of this kind are not unique to theoretical reason: they may also be used to avoid self-contradiction in the practical use of reason. To begin with, they are instantiated in the first formulation of the categorical imperative: one should act only according to that maxim whereby one can, at the same time and without contradiction, will that it should become a universal law (cf., e.g., MAN, AA IV, 421, 07-08; for the first formulation of the categorical imperative as a TE, see Illies 2007, 313). Secondly, Kant talks explicitly about an "experiment" of

practical reason in a passage of the *Critique of Practical Reason.* Exactly as he does in the *Critique of Pure Reason,* Kant remarks that philosophy has an additional problem with respect to natural science or mathematics, since a philosopher cannot use intuition as a foundation but, "almost like a chemist" *(beinahe wie der Chemist),* he or she can at any time conduct "an experiment with every man's practical reason" identifying the relevant moral (pure) principle of determination. The chemist adds alkali to a solution of lime in hydrochloric acid, in order to get the lime to precipitate. Similarly, we may add, even if "only in thought" *(nur in Gedanken),* the moral law (as a determining principle) to the empirically affected will, in order to avoid a contradiction of reason with itself (cf. *Kritik der praktischen Vernunft,* AA, V, 92-3, lines 33-7 and 1-10). Finally, Kant uses the expression "experiment of pure reason" also in the essay "On Eternal Peace." By abstracting all empirical matters (such as the various relationships between citizens or between states) from the idea of public law, Kant extends the first formulation of the categorical imperative to the definition of public law. How may we judge whether the "form" of the law is realized in any particular case? By means of "an experiment of pure reason," which is "a criterion which is easy to use in experience and is presented a priori in reason." We have to judge whether the "form" of the law can or cannot be combined with the principles of a moral agent. If not, we can "immediately recognize the falsity (illegality) of a legal claim *(praetensio juris)" (Zum ewigen Frieden. Ein philosophischer Entwurf* AA, VIII 381, 04-18).

To the extent that experiments of pure reason are methodologically similar to scientific experiments, one may agree with Shi-Hyong (2008) and Fulkerson-Smith (2013) that "experiments of pure reason" show Bacon's influence on Kant. However, one should not stress similarities at the expense of fundamental differences between the two kinds of experimentation. It is true that Kant associates reason's capacity to abstract the material from the formal (a priori) components of a phenomenon with the chemist's activity of experimenting. However, it is the use of experiments of pure reason that distinguishes transcendental philosophy from the empirical sciences and mathematics. For this reason we cannot accept Westphal's claim that transcendental arguments have the status of psychological, "epistemic reflections," which attempt to find out "whether Kant's inventory of cognitive capacities holds true *of us"* (Westphal 2003, 141) "by using wildly counter-factual TEs" (Westphal 2004, introduction; see also 18-34). Kant's idea should be developed such that it reconciles a difference between scientific and philosophical TEs with a methodological naturalism that admits no difference in kind between the methods of science and philosophy (see Buzzoni 2016).

**3 Novalis', Lichtenberg's, and Orsted's concept of TE**

Among the authors cited as precursors of neo-Kantian accounts of TEs there are Novalis (1772 —1801) and Georg Christoph Lichtenberg (1742-1799) (cf. Schöne 1982; Daiber 2001, 24; Schildknecht 1990, 123-69; Kühne 2005, 220-21; Fehige 2012; Fehige and Stuart 2014, 181-98).

Though we find in Novalis the notions both of an "experimental physics of the mind" and of "complete experiments" as a synthesis of physical experiments and poems of productive

imagination (cf. Fehige 2012, 156), his importance in the history of the Kantian approaches to TEs should not be overestimated. First, a generic Kantian influence is insufficient to qualify one as a "Neo-Kantian" (cf. Anderson 2005, 289). Secondly, this influence in Novalis is overlaid with many others (Fichte, Schelling, Spinoza, Plotinus, Plato, Schiller, Goethe, Böhme, Baader, etc.). Thirdly, scientific experiments for Novalis are no more reliable than philosophical or fictional ones — a thesis that blurs, in full accord with German *Naturphilosophie* and Romanticism, Kant's distinction between physics, mathematics and philosophy.

More important than Novalis was Lichtenberg, who spoke of "experimenting with thoughts" ("mit Gedanken [...] experimentieren") and of *"experimenting* with ideas" ("mit Ideen *experimentieren")* (Lichtenberg 1968 [1994], K308). Lichtenberg is also well aware that TEs share with REs the character of questions put to nature (Lichtenberg 1971 [1994], K 308 and 310). Moreover, he maintains that the careful use of the subjunctive mode may be used experimentally to break apart conceptual connections that limit scientific creativity (cf. Fehige and Stuart 2014, 187). Finally, Lichtenberg notes that TEs require hypothetical idealizations (cf. Lichtenberg 1971 [1994], K 311) which consist in carrying to the limit the value of a variable (Lichtenberg 1971 [1994], J 1644) — an idea later to be taken up by Mach.

The first theory of TEs that can be defined as Kantian or neo-Kantian in a more rigorous sense was developed by Hans Christian Ørsted. Ørsted adopts Kant's idea of scientific experiment: "To make experiments is to ask questions of nature, but no one can do this usefully unless he knows what to ask about" (Ørsted 1811 [1920], §14. English translation, 294). However, Ørsted criticizes Kant for not considering chemistry as a proper natural science because it was not amenable to mathematization: TE is for Ørsted a particular a priori method for studying dynamic chemical processes (cf. Witt-Hansen 1976, 53-4; Kühne 2005, 92-165; Cohnitz 2008; Buzzoni 2011, 97-99; Fehige 2012, §4; Wilson 1998 [2014], XXII—XXIII; Christensen 2013, 267).

According to Ørsted, TEs are widely used in mathematics, where they consist in the construction of concepts in the intuition: "When we allow a point to move in our imagination in order to produce a line, or a line to rotate around one of its *extremities* to describe a circle with the other, what is this but a TE? Differential and integral calculus consist of nothing but such thought experiments and considerations of them" (Ørsted 1811 [1920], English translation, 296). But TEs are not only demonstrations in the sense of Kant's *Critique of Pure Reason,* which have no "point of contact with today's use of the expression in philosophy of science" (Cohnitz 2008, 407-8). As I said earlier, TEs constitute an a priori method that can be extended to any field of the natural sciences, including chemistry (cf. Ørsted 1822, 482-3; cf. also Fehige 2012, 151-52, who provides a detailed criticism of Cohnitz's point of view).

Kühne 2005 argued that Ørsted wavers between two positions: on the one hand he claims that TEs are only a necessary precondition of empirical science, which is always ultimately founded on REs (cf. Ørsted 1811, §19; Christensen 2013, 271); on the other hand he ascribes to TEs the capacity of determining some fundamental laws of natural science with a degree of certainty in principle identical with that of mathematics (cf. Ørsted 1851, 7-8, and Ørsted 1850/1851, vol. 2 (1851), §§14 and 15; Kühne 2005, respectively 138 and 149-50; Wilson 1998 [2014] ).

This wavering depends upon the fact that the status of Kant's a priori is ambiguous and open to opposing interpretations. As is well-known, according to Kant, (1) all "a priori knowledge" is independent of experience (KrV B 3, AA 28, B 117, 269, AA 99-100, 187); (2) the a priori is the condition upon which all experience depends (KrV B 269, AA 188); and (3) the distinctive traits of the a priori are "unconditional necessity" *(unbedingte Notwendigkeit)* and "true," *(wahre)* "strict," *(strenge)* or "absolute universality" *(absolute Allgemeinheit)* (cf. KrV B 3-5, AA 28-30, B 64, AA 68). However, Kant oscillates between two distinct and ultimately contradictory interpretations of the a priori thus defined, one functional and one material. On the one hand, Kant says that the a priori determines only the formal conditions of experience and is devoid of material content; on the other hand, he ascribes content to the (synthetic) a priori since he believes that his transcendental philosophy provides a genuine metaphysical foundation for Newton's natural philosophy, and that part of Newton's *Principia* (its "pure part") provides us with universal and necessary knowledge in need of no further substantial modification (cf. Friedman 1992, 4).

Ørsted's account of TEs, under the influence of German romantic *Naturphilosophie,* develops Kant's a priori in the second (material) rather than the first (functional) sense. In this respect, it is important to note that Ørsted's method of TE is reminiscent of Kant's so-called "transition project" (published posthumously), concerning a possible transition from the special metaphysics of nature contained in the *Metaphysical Foundations* to physics itself (Christensen 2013, 267; on Kant's "transition project," cf. also Hall 2009).

As I stated above, Ørsted's point of view had no real influence on the historical development of the concept of TE. A glance at the history of the philosophy of science in the light of the previous considerations is sufficient to explain this fact. According to Moritz Schlick, the question "Is there a material a priori?" must be answered in the negative: "The empiricism which I hold believes itself to know with certainty that, as a matter of principle, all statements are either synthetic a *posteriori* or tautological; according to it, synthetic a *priori* statements seem to be a logical impossibility" (Schlick 1932 [1969], 25). This rejection of the material a priori was ultimately accepted not only by the Logical Empiricists but also by principal exponents of the philosophy of science, since the birth of the discipline at the end of the nineteenth century. The material a priori was rejected either in the empiricist spirit of Mach, Neo-positivism, Bridgman, and Popper, or in the conventionalist spirit of Poincare and Duhem, later taken up by the relativist philosophy of science of the 1960s and the "sociological turn," which construed the a priori as a changeable function of historically shifting pragmatic interests. Kant's a priori was therefore regarded as an error, perhaps as an inevitable one before Riemann, Einstein, and the founders of quantum physics, but nevertheless as an error (cf., for example, Mach 1883 [1933], 458-9; Poincaré 1902 [1968], 74-5; Reichenbach 1920, 1-5, and 1938, 12 and 346; Einstein 1924, 1688-9; Bridgman 1927, 3-9; Popper 1963, ch. 7).

It was therefore no accident that broadly Kantian approaches to TEs grew out of concepts of the a priori that were essentially different from Ørsted's. I myself view the a priori as universal but purely formal or functional, that is, devoid of any content. Yiftach Fehige views the a priori as material, but contingent and relativized. In both cases we find the same rejection of an

essentially idealistic notion of the a priori in the sense that was fully developed by Hegel. We shall devote the rest of this chapter to examining these two neo-Kantian approaches to TEs.

4 **TEs from a neo-Kantian point of view: the transcendental-operational account**

After Ørsted, my own transcendental-operational account was the first "Kantian treatment" of TEs (Brown 2011, 202). As Lichtenberg and Ørsted did, I began with Kant's definition of "experiment," according to which a RE is a "question" put to nature; however, I interpreted the notion of experiment from the viewpoint of an a priori that is devoid of any particular content. It is true that Kant ascribes some content to the (synthetic) a priori, but only a functional reading of the a priori agrees with the spirit of Kant's philosophy and can be used for developing a consistent account of TEs. In fact, there is in Kant a fundamental tendency to consider the synthetic a priori in a purely formal (or rather, *functional)* sense which he contrasts with the *material* conditions, which are given through sensation. For example, Kant emphasized that categories not applied to sense content are "merely functions of the understanding for concepts" and "cannot ... be employed in any manner whatsoever," whether empirical or transcendental (cf. respectively KrV B 187, AA 139, B 305, AA III 208). It is in connection with this sense of the a priori that Kant claims that the manifold is given through the "I think" as the supreme condition of the possibility of experience: the "I think" cannot have a content of its own and must be conceived as a mere form or function (KrV B §16, AA 110).

On the basis of this concept of the a priori, I developed an account that mediates between Brown and Norton by considering TEs from two distinct but connected points of view: transcendental and operational (Buzzoni 2004, 2007, 2008, 2011). On the one hand, the account is transcendental in a very usual sense of the word (cf. Hatfield 1990, 79): TEs are the condition of the possibility of REs because, without the a priori capacity of the mind to reason counterfactually, we could not devise any hypothesis and would be unable to plan the corresponding RE that should test it. This capacity underpins the distinction in principle - a properly transcendental distinction - between TEs and REs. This distinction "cannot be suppressed, since it is the same distinction between the hypothetical-reflexive domain of the mind (which can always contradict itself) and reality (which can always occur and develop in only one way)" (Buzzoni 2013a, 99-100).

But while there is a distinction in principle between TE and RE on the transcendental level of the a priori capacity of the mind to reason counterfactually, the exclusively functional character of the a priori entails that, on the operational-methodological level, TEs - considering their empirical content and actual performance within a discipline - are and function essentially like REs (cf. Buzzoni 2013a, 98).

At the operational-methodological level, an empirical TE "anticipates ... a hypothetical experimental situation so that, on the basis of previous knowledge, we are confident that certain interventions on the experimental apparatus will modify some of its aspects (or `variables') with

such a degree of probability that the actual execution of the experiment becomes superfluous." (Buzzoni 2013a, 97, italics dropped; cf. also 2008, 93) Now, how ought we to interpret this "previous knowledge" in the light of a functional, non-hypostatized a priori? If pure reason cannot provide any content, the entire content of empirical TEs must come in principle from empirical-operational interventions on reality, that is, from actual experimentation. In this sense, a TE would be devoid of empirical meaning (that is, it would not be a TE proper to empirical science) if it did not possess, even while it is still in our minds, an (at least implicit) reference to experience.

The operational-methodological similarity and the transcendental distinction of TEs and REs are best illustrated by Kant's example of a hundred dollars (cf. KrV B 627, AA III 401).

On the one hand, from the perspective of the analysis of the intensions of the respective concepts, our conception of a hundred dollars remains the same irrespective of whether I own them or not. In the same sense, every (empirical) TE corresponds to a RE that satisfies the same conceptual characteristics, and vice versa. All REs may also be thought of as realizations of TEs; conversely, all empirical TEs must be conceivable as preparing and anticipating REs: They must, that is, anticipate a connection between objects which, when thought of as realized, makes the TE coincide completely with the corresponding RE (cf. Buzzoni 2008, 110). For this reason, TEs have the same operational-methodological traits as REs: both are constituted by a theory and a particular, well-specified experimental apparatus (Buzzoni 2013a, 97-8); both ask questions about nature and its laws; both apply Mach's method of variation, whereby some variables are systematically modified to establish which relation of dependence, if any, holds between them; both do that in a theory-laden and idealized way, so that the meaning of both must always be interpreted; both exemplify an "inductive-experimental use of reason" (Buzzoni 2008, 69 and 109; more recently the same claim is made by Clatterbuck 2013); in both cases visualization, perspicuity, intuitive appeal, and clarity are important because empirical TEs apply hypotheses to specific particular cases (those relevant for testing their truth or falsity).

On the other hand, from the transcendental perspective, thought dollars, like TEs, exist only in the sphere of the possible, while real dollars, like REs, occupy a specific place among the interactions between our bodies and the surrounding reality (cf. Buzzoni 2008, 113).

An important consequence of this unity and distinction of REs and TEs is that neither would be what it is outside their mutual relationship. On the one hand, despite their conditioned empirical power, TEs are very important in science for the following reasons: (1) at the most fundamental (transcendental) level, without TEs there would not be REs (we would not know how to put our questions to nature); (2) at the operational-methodological level, TEs, to the extent that they are based on well-established scientific facts and laws of nature, even if they are not realized or we decide not to realize them, can "inductively extend our knowledge" (Buzzoni 2008, 96) and support, at least provisionally, fresh scientific claims. On the other hand, without already realized TEs there would be no reliable empirical laws on which new TEs can base their anticipated answers to new questions. To sum up, they are complementary in a typical Kantian sense: "(empirical) TEs without REs are empty; REs without TEs are blind" (Buzzoni 2013a, 100).

A number of objections have been raised against my account. Brown and Fehige 2014 have objected that, "given the many scientific TEs that cannot be realized in the real-world, Buzzoni might be conflating TEs with imagined experiments to be carried out in the real-world." (cf. also Fehige 2013, 62-3). Now, is an imagined (real) experiment a TE? Yes and no. Surely not, if it is considered apart from any intentional planning. On the other hand, an imagined (real) experiment is a TE if mental "images" are thought of as mental *projects* with experimental value, that is, if they are seen in connection with the ability to reason about the outcomes of a series of possible actions. Remembering or imagining a hundred dollars is not an empirical TE. But this imagining becomes a TE if we place it in a planning context: for example, if we see it from the standpoint of a forger who anticipates in thought how to illegally produce the money (maybe without putting his project into practice because of technical difficulties or because it is too risky).

Another important objection is that for many TEs the requirements of empirical realizability and technical-operational testability appear to be either in principle unsatisfiable or superfluous (Fehige 2013). I have replied to this objection by insisting on the difference between *de facto* and *in principle* realizability. Obviously, there are good empirical TEs that, *de facto,* neither are nor will ever be REs: limitations of our intelligence or resources may indefinitely postpone the realization of a TE, perhaps certain technical barriers will remain forever unsurmountable, and so on. More than this: many TEs are as important as they are exactly because we are either technically incapable of realizing them or unwilling to carry out the corresponding REs, even though we have good reasons to regard them as in principle realizable. The crucial point is that no empirical TE can ever be absolutely unrealizable without losing its empirical character. We believe that empirical TEs are scientifically useful and reliable because we presuppose that, if they were realized, the sequence of events *that they describe according to causal connections which we assume to be operative in the real world,* would occur in the way they anticipate, and would lead to the consequences that they predict. This holds in principle, no matter how remote the realizability of certain TEs may be. If this assumption is abandoned, TEs cease to be scientific in any important sense. A narrative TE can be incompatible with reality without this having any immediate implications for its truth. On the contrary, if we suppose that Einstein's lift scenario is physically impossible not just in practice, but in principle, it would then no longer be directly relevant to science. Without the assumption of the possibility of a practical realization, we abandon science and are left with only the play of imagination (cf. Buzzoni 2013b, 280-1).

A third and final objection is that I do not give a criterion for distinguishing good from bad TEs independently of physical experiments (Fehige 2013, 63). My reply is that, while it is true that in cases of doubt we must turn to real experiments, which remain the ultimate criterion for all empirical TEs, the general criterion of realizability is, on its own, insufficient to distinguish good from bad empirical TEs. In order to do that, the general criterion of realizability must be translated into concrete and detailed criteria and methods, which are as numerous as the different problems of human life itself. Since the number of these problems is indeterminate, no complete list of criteria or methods can be given a *priori* or (what comes to the same thing) they cannot be reduced to a single criterion or method. For example, we may examine whether a TE

only uses widely accepted empirical generalizations, unquestionable assumptions, or everyday abilities so that the carrying out of the corresponding RE could only lead to a foregone conclusion. Otherwise we may detect ambiguities in the meaning of the question that the TE puts to nature. It may also be possible to investigate critically whether a TE rests upon or is consistent with inductively well-supported laws (provided these laws are relevant to the case at hand). Or we may resort to indirect means of assessing a TE, i.e., to methods which are only indirectly related to its realizability in principle, for example, by asking questions such as, "Does the TE answer an important question or does it sidetrack thought into collateral issues?" "Is it fruitful, for example in the sense of opening up new prospects for research?"

It is only through a process of interpretation and reinterpretation of the past history of knowledge (in this case, of the past evaluations of TEs) that we know anything about such methods and criteria (this process is an aspect of the hermeneutic aspect of TEs: cf. Buzzoni 2008, ch. 3 §3). Only by reconstructing past evaluations of TEs can one discover criteria that form a body of fairly stable methods for the attainment of good evaluations. These methods are, in the last analysis, authorized by past experience and successful REs, but they are not a fixed and unchangeable catalogue: the evaluator applies them to new cases and transforms them so that the discussion about the set of good methods never ends (cf. Buzzoni 2013b).

Developing some points of my view, Mike Stuart has tried to interpret the a priori in the light of Kant's notion of transcendental "schemata" which connect the senses with the understanding. Stuart claims that for Kant, the schemata are not concepts themselves, nor rules for application of concepts (Schaper 1964, 272). They are the action or performance of connecting concepts to experience or other concepts. Applying and connecting concepts to real-world activity requires exercising the faculty of judgment, which cannot be taught via rules, but has to be practiced and improved, usually by the use of examples (cf. Allison 1983 [2004], 207). Adopting my notion of the a priori, which relies on the capacity of the mind to consider things counterfactually, Stuart contends that TEs can fix the meaning of difficult theoretical structures (principles, equations, concepts, models, etc.) through the use of imaginary examples. This increases understanding for Stuart, not knowledge. He emphasizes that while understanding is an important step on the path to new knowledge, it is not new knowledge itself. Therefore on the one hand, Stuart sees imagination as the *necessary* a priori condition of (meaningful) "understanding" alone, not of (true) "knowledge": the bare use of the imagination in science cannot be justified by cognitive science or by philosophy. It is only justified in a transcendental sense because it is always necessarily presupposed by both. That is why it is a *sine qua non* of scientific understanding" (Stuart 2017). And on the other hand, this idea is developed from the perspective of cognitive science. In accordance with the top-down processing accepted by many cognitive scientists today, TEs can connect theoretical structures to experience or other theoretical structures that are already meaningful. Stuart agrees with me that at bottom, scientific knowledge and understanding result from operationalization. However, in difficult cases, as in the development of quantum mechanics, the connection between partially operationalized structures must be made via the imagination.

**5 TEs from a neo-Kantian point of view: the relativized a priori account**

As we have seen, under the influence of German romantic *Naturphilosophie* Ørsted's account of TEs developed Kant's a priori in a sense that may be called material and that was rejected by later schools of philosophy of science. I moved in the opposite direction by retaining the universality and necessity of Kant's a priori, while subtracting from it any particular content, which instead comes only from experience. On the contrary, Yiftach Fehige provided a third way of defending a neo-Kantian approach to TEs: under the influence of Michael Friedman, Fehige retained the material character of the a priori, while abandoning its necessity and universality (Fehige 2013).

Reichenbach distinguished two meanings of Kant's notion of a priori knowledge: "Kant's concept of a priori has two different meanings. First, it means “apodictically true” *(apodiktisch gültig),* “true for all times” and secondly, “constituting the concept of object” (Reichenbach 1920 [1965], 46. English translation slightly modified). Many authors followed Reichenbach by rejecting the first meaning while retaining the second. The result is "a relativized and dynamical conception" of the Kantian a priori: a priori principles are not universal and necessary (and therefore fixed for all times), but they are preconditions for knowledge, in the sense that they are constitutive of the objects of knowledge according to a given theory, with respect to which they are thus relative. A priori principles may provide an answer to Kuhn's challenge because they guarantee the unity of science "across revolutionary paradigm-shifts" (Friedman 2002, 185; cf. also Friedman 2010, 716; on the "relativized a priori" see, for example, Philström and Sitonen 2006; Richardson 2010, 282; Tanona 2010, 423-4; De Boer 2010).

The idea in itself is not new. Not only Poincaré and Reichenbach (cf., for example, Friedman 2002, 173-5), but also philosophers as different as Lorenz (1941/42, 1959, 1973 [1977]), Wittgenstein, Popper, Kuhn, Miller (1941), Pap (1944 and 1946), Milmed (1961), Scheffler (1967), Körner (1969 [1979]), and perhaps also Lewis and Dewey (cf. Stump 2011, 247) shared a view of this kind. What is new is the application of the relativized a priori to the understanding of TEs. This notion assumes a content for the a priori that can be made explicit through TEs, without suffering under Ørsted's delusion that TEs are capable of establishing some kind of necessity with respect to empirical statements concerning nature. In fact, this concept of the a priori has much less to do with Husserl's material a priori (cf., for example, Husserl 1950, I, §§9 and 16), which was criticized by Schlick, than with the conception - to mention only one of many possible examples - of Arthur Pap (cf. 1944 and 1946), who distinguished between a "material," a "functional," and a "formal" (or analytic) a priori. In this way, the material aspect of Kant's a priori, which for me was a limit that needed to be overcome, becomes for Fehige a point of departure for developing an alternative neo-Kantian account of TEs. More precisely, according to Fehige the kind of synthetic a priori knowledge that Kant ascribed to the "pure part" of natural science is the strongest reason to return to Kant in order to find a middle ground between Platonism and empiricism: a Kantian theory of TEs dispenses with Platonic entities and intellectual intuition of any kind, while taking seriously "the idea that TEs can reveal

information about the world that goes beyond pure empirical input" (Fehige and Stuart 2014, 206; cf., in the same vein, see Fehige 2012, 147).

For Fehige, this neo-Kantian account is an extension and refinement of Kuhn's views about TEs. It must be admitted that the challenge to the analytic-synthetic dichotomy inherent in Kuhn's idea that TEs teach us both about our concepts and about the world, clearly moves in the direction of some sort of synthetic a priori and posits an intrinsic connection between real and TEs. Moreover, Kuhn did not try to overcome the separation between analytic and synthetic judgements at the transcendental level, but rather at the historical-empirical level of the relativity of Wittgenstein's language games (Buzzoni 2005).

Fehige is therefore right to see the relationship between his own views and those of Gendler (2000) and McAllister (1996, 2004), who have also taken up Kuhn's point of view. Gendler's theory is a development of Kuhn's conception in the sense that TEs lead to new knowledge of the world by reconfiguring and revising pre-existing epistemological meta-frameworks. Similarly, James McAllister argued that the Galileo's TE against Aristotle's theory of motion, which involves dropping cannon balls of different weights, could not have been of intrinsic evidential significance without important changes in the epistemological meta-framework. In any case, it would be impossible in principle to deduce a contradiction from a theory, thus falsifying it, if there were no shared epistemological meta-framework in the sense of a relativized a priori (cf. Fehige 2013, 69-71).

From the point of view of a relativized and contingent a priori, Fehige develops a straightforward and plausible answer to the main epistemological challenge raised by TEs, namely how can TEs, which do not rely on new material drawn from experience (as REs do), lead to unexpected conclusions sometimes capable of casting doubt on well-confirmed empirical theories. The relevant new information that comes through "reconceptualization" is not empirical, but "philosophical and relates to the Kantian a priori ä la Friedman ... It is a reconfiguration of internal conceptual space which depends on changes in the epistemological meta-framework that explains the progress on an empirical matter ... without new empirical information about the world" (Fehige 2013, 69).

Moving from this idea of the a priori, Fehige also provides a plausible answer to the problem of finding criteria for assessing the quality of a TE. Einstein's clock-in-the-box TE suffers from an important shortcoming in comparison with Galileo's falling bodies TE. While Einstein's TE "does not establish sufficiently a link between the Kantian a priori à la Friedman and the empirical matter under consideration," Galileo's TE "is successful in that it moves even the Aristotelian to accept the changes in the Kantian a priori enabling new conceptual possibilities which help the Aristotelian to perceive a familiar anomaly in a new light" (Fehige 2013, 71).

The neo-Kantian account of TEs based on the contingent a priori raises two questions. The first one relates to a theoretical rather than to a historical context, and concerns the tenability of Friedman's answer to Kuhn's challenge of incommensurability and relativism across revolutions. The second question is to what extent the account of TEs based on the contingent a priori can be said to be truly Kantian. As to the first issue, it has been observed that Friedman's neo-Kantian account ascribes to changeable principles a role that they cannot have (cf. Howard 2010) and that it is incapable of solving the main problem which it was intended to solve, that is

the problem of incommensurability between paradigms in Kuhn's sense (Buzzoni 2013b, 289-91). Unfortunately I lack the space to address this controversial issue here (but see, for example, Boghossian and Peacocke 2000).

As far as the second question is concerned, we certainly have to preserve the spirit of Kant's philosophy more than its letter. But the question arises whether we can obey the spirit of Kant's philosophy without obeying considerable parts of its letter as well. First, to say that Friedman's a priori is Kantian because it is material in character and has the function of making experience possible seems to allow too many positions to be called "Kantian" (Buzzoni 2013b, 286-91). Secondly, although Kant's a priori is not entirely free from material elements (and in this sense it has something in common with Friedman's a priori), the utility of turning it into something that has all the properties found in the empiricist concept of experience — relativity, variability, particularity — is very doubtful. Some authors have argued that, if Kant's a priori is understood as phylogenetically or psychologically a posteriori, the independence from experience, which is the main characteristic of Kant's a priori, is lost (Kuhle and Kuhle 2003, 214; cf. also Haller 1987, 19-20).

Whatever answer one may give to these questions, which must here remain open, the current state of discussion shows that the relativized a priori account is likely to remain an important reference point in the debate about TEs: all the more important, to the extent that the conception of the contingent a priori is a very widespread one today.

**6 Conclusion**

We have outlined in this contribution a critical history of the concept of TE from a broadly Kantian viewpoint. The thread of our discussion has been Kant's position on the nature of the a priori and how each neo-Kantian theory of TEs can be understood in terms of its own position on this issue. Section 2 examined some aspects of Kant's philosophy that are related to today's debate on TEs. Section 3 was devoted to Ørsted's first important Kantian theory of TE, which had no effective influence on the historical development of the concept because his commitment to romantic *Naturphilosophie* led him to see TE as a special method for an a priori physics.

Given the rejection of a material a priori by the most important schools of philosophy of science, it was no accident that a broadly Kantian approach to TEs was recently taken up by moving to quite different interpretations of the a priori, namely as universal and necessary, but devoid of content (myself and Mike Stuart), and as a set of principles which are taken as constitutive of experience, but only in a contingent and relativized way (Yiftach Fehige).

**References**

Allison, H. (1983 [2004]) *Kant's Transcendental Idealism,* New Haven: Yale University Press.

Anderson, R. L. (2005) "Neo-Kantianism and the roots of anti-psychologism," *British journal for the History of Philosophy* 13: 287-323.

Boghossian, P. and Peacocke, C. (2000) *New Essays on the A Priori,* Oxford: Oxford University Press.

Bridgman, W. (1927) *The Logic of Modern Physics,* New York: MacMillan.

Brown, J. R. (1991 [2011]) *Laboratory of the Mind: Thought Experiments in the Natural Sciences,* London: Routledge (quotations are from the 2nd ed., 2011).

Brown, J. R. and Fehige, Y. (2014) "Thought experiments," *Stanford Online Encyclopedia of Philosophy.* Available here: http://plato.stanford.edu/entries/thought-experiment/ (accessed 15 November 2016).

Buzzoni, M. (2004) *Esperimento ed esperimento mentale,* Milan: Angeli.

Buzzoni, M. (2005) "Kuhn und Wittgenstein: Paradigmen, Sprachspiele und Wissenschaftsgeschichte," in *Zeit und Geschichte/Time and History,* edited by F. Stadler and M. Stöltzner, Kirchberg: Ludwig Wittgenstein Gesellschaft.

Buzzoni, M. (2007) "Zum Verhältnis zwischen Experiment und Gedankenexperiment in den Naturwissenschaften," *Journal for General Philosophy of Science* 38: 219-237.

Buzzoni, M. (2008) *Thought Experiment in the Natural Sciences: An Operational and Reflexive-Transcendental Conception,* Würzburg: Königshausen and Neumann.

Buzzoni, M. (2011) "Kant und das Gedankenexperiment," *Deutsche Zeitschrift für Philosophie* 59: 93-107.

Buzzoni, M. (2013a) "Thought experiments from a Kantian point of view," in *Thought Experiments in Philosophy, Science, and the Arts,* edited by M. Frappier, L. Meynell, and J. R. Brown, London: Routledge.

Buzzoni, M. (2013b) "On thought experiments and the Kantian a priori in the natural sciences: A reply to Yiftach J. H. Fehige," *Epistemologia* 36: 277-293.

Buzzoni, M. (2016) " Thought experiments in philosophy: A neo-Kantian and Experimentalist point of view," *Topoi,* DOI 10.1007/s11245-016-9436-6

Christensen, D. Ch. (2013) *Hans Christian Ørsted:: Reading Nature's Mind,* Oxford: Oxford University Press.

Clatterbuck, H. (2013) "The epistemology of thought experiments: A non-eliminativist, non-platonic account," *European journal for Philosophy of Science* 3: 309-329.

Cohnitz, D. (2008) "Ørsteds `Gedankenexperiment': eine Kantianische Fundierung der Infinitesimalrechnung?" *Kant-Studien* 99: 407-433.

Daiber, J. (2001) *Experimentalphysik des Geistes. Novalis und das romantische Experiment,* Göttingen: Vandenhoeck and Ruprecht.

De Boer, K. (2010) "Kant, Reichenbach, and the fate of a priori principles," *European journal of Philosophy* 19: 507-531.

Einstein, A. (1924) "Review of Elsbach, Alfred. Kant und Einstein. Untersuchungen über das Verhältnis der modernen Erkenntnistheorie zur Relativitätstheorie (Berlin-Leipzig: de Gruyter)," *Deutsche Literaturzeitung* 45: 1688-1689.

Fehige, Y. (2012) "`Experiments of pure reason': Kantianism and thought experiments in science," *Epistemologia* 35: 141-160.

Fehige, Y. (2013) "The relativized a priori and the laboratory of the mind: Towards a neo-Kantian account of thought experiments in science," *Epistemologia* 36: 55-73.

Fehige, Y. and Stuart, M.T. (2014) "On the origins of the philosophy of thought experiments: The forerun," *Perspectives on Science* 22: 179-220.

Friedman, M. (1992) *Kant and the Exact Sciences,* Cambridge: Harvard University Press.

Friedman, M. (2002) "Kant, Kuhn, and the rationality of science," *Philosophy of Science* 69: 171-190.

Friedman, M. (2010) "Synthetic history reconsidered," in *Discourse on a New Method: Reinvigorating the Marriage of History and Philosophy of Science,* edited by M. Domski and M. Dickson, Chicago: Open Court.

Fulkerson-Smith, B. (2013) "Kant's illuminating experiment: On the placement, purpose and essential procedure of the experiment of pure reason in the critique of pure reason," *Society and Politics* 7: 62-83.

Gendler, T. S. (2000) *Thought Experiment: On the Power and Limits of Imaginary Cases,* London: Garland.

Hall, B. (2009) "`Effecting a transition': How to fill the gap in Kant's system of critical philosophy," *Kant-Studien* 100: 187-211.

Haller, R. (1987) "Kommentar zu dem Vortrag von Konrad Lorenz über Evolution und Apriori," in *Die evolutionäre Erkenntnistheorie. Bedingungen-Lösungen-Kontroversen,* edited by R. Riedl and F. Wuketits, Berlin: Parey.

Hatfield, G. (1990) *The Natural and the Normative: Theories of Spatial Perception from Kant to Helmholtz,* Cambridge: MIT Press.

Howard, D. (2010) "`Let me briefly indicate why I do not find this standpoint natural': Einstein, General Relativity, and the contingent a priori," in *Discourse on a New Method: Reinvigorating the Marriage of History and Philosophy of Science,* edited by M. Domski and M. Dickson, Chicago: Open Court.

Husserl, E. (1950) *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie,* Den Haag: Martinus Nijhoff.

Kalin, M. G. (1972) "Kant's transcendental arguments as Gedankenexperimente," *Kant-Studien* 63: 289-328.

Körner, S. (1969 [1979]) *Fundamental Questions of Philosophy: One Philosopher's Answers,* Brighton, UK: The Harvester Press; fourth edition 1979: Atlantic Highlands, NJ: Humanities Press.

Kuhle, M. and Kuhle, S. (2003) "Kants Lehre vom Apriori in ihrem Verhältnis zu Darwins Evolutionstheorie," *Kant-Studien* 94: 220-239.

Kühne, U. (2005) *Die Methode des Gedankenexperiments,* Frankfurt: Suhrkamp.

Lichtenberg, G. Ch. (1968 [1994]) *Schriften und Briefe. Erster Band: Sudelbücher I,* edited by W. Promies, Wien: Carl Hanser. Aphorisms are quoted by the letter assigned to a book, namely

A—L, and the number of the aphorism within that book. Quotations are from the 1994 edition.

Kühne, U. (2005) *Die Methode des Gedankenexperiments,* Frankfurt: Suhrkamp.

Lichtenberg, G. Ch. (1971 [1994]) *Schriften und Briefe. Zweiter Band: Sudelbücher I. Materialhefte, Tagebücher,* edited by W. Promies, Wien: Carl Hanser. Aphorisms are quoted by the letter assigned to a book, namely A—L, and the number of the aphorism within that book. Quotations are from the fifth 1994 edition.

Lorenz, K. (1941-1942) "Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie," *Blätter für deutsche Philosophie* 15: 94-125.

Lorenz, K. (1959) "Gestaltwahrnehmung als Quelle Wissenshaftlicher Erkenntnis," *Zeitschrift für experimentelle und angewandte Psychologie* 6: 118-165.

Lorenz, K. (1973 [1977]) *Die Rückseite des Spiegels,* München: Piper. English translation: *Behind the Mirror,* London: Methuen.

Mach, E. (1883 [1933]) *Die Mechanik in ihrer Entwickelung. Historisch-kritisch dargestellt,* Leipzig: Brockhaus (quotations are from the 9th ed., 1933).

McAllister, J. W. (1996) "The evidential significance of thought experiments in science," *Studies in History and Philosophy of Science* 27: 233-250.

McAllister, J. W. (2004) "Thought experiments and the belief in phenomena," *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 71: 1164-1175.

Miller, D. L. (1941) "The a priori in contemporary thought," *Philosophy of Science* 8: 20-25.

Milmed, B. K. (1961) *Kant and Current Philosophical Issues: Some Modern Developments of His Theory of Knowledge,* New York: New York University Press.

Ørsted, H. C. (1811 [1920]) " “Første Indledning til den almindelige Naturlaere,” in Naturvidenskabelige Skrifter, vols I–III, edited by K. Meier, Copenhagen: Host and Son, 1920, 151-190. English translation: "First introduction to general physics: A prospectus of lectures in this science," in *Selected Scientific Works of Hans Christian Ørsted,* 282-309, translated and edited by K. Jelved, A. D. Jackson, and 0. Knudsen, Princeton: Princeton University Press.

Ørsted, H. C. (1822) "Über Geist und Studium der allgemeinen Naturlehre," *Gehlens journal für Chemie und Physik* 36: 458-488.

Ørsted, H. C. (1850-1851) *Der Geist in der Natur,* 2 vols, Munich: Cotta. Translated by L. Homer and J. B. Homer, London: Bohn, 1952.

Ørsted, H. C. (1851) *Der mechanische Theil der Naturlehre,* Braunschweig: Vieweg und Sohn. Pap, A. (1944) "The different kinds of a priori," *Philosophical Review* 13: 465-484.

Pap, A. (1946) *The A Priori in Physical Theory,* New York: King's Crown.

Philström, S. and Sitonen, A. (2006) "The transcendental method and (post) empiricist philosophy of science," *Journal for General Philosophy of Science* 36: 81-106.

Poincare, H. (1902 [1968]) *La Science et l'Hypothese,* Paris: Flammarion (quotations are from the 1968 edition).

Popper, K. R. (1963 [1972]) *Conjectures and Refutations: The Growth of Scientific Knowledge,* 4th ed., London: Routledge and Kegan Paul.

Reichenbach, H. (1920 [1965]). *Relativitätstheorie und Erkenntnis a priori,* Berlin: Springer. English translation by M. Reichenbach, *The Theory of Relativity and A Priori Knowledge,* Berkeley: University of California Press.

Reichenbach, H. (1938) *Experience and Prediction,* Chicago: University of Chicago Press.

Richardson, A. (2010) "Ernst Cassirer and Michael Friedman: Kantian or Hegelian dynamics of reason?" in *Discourse on a New Method: Reinvigorating the Marriage of History and Philosophy of Science,* edited by M. Domski and M. Dickson, Chicago: Open Court.

Schaper, E. (1964) "Kant's schematism reconsidered," *The Review of Metaphysics* 18: 267-292. Scheffler, I. (1967) *Science and Subjectivity,* Indianapolis: Bobbs-Merrill.

Schildknecht, C. (1990) *Philosophische Masken: Literarische Formen der Philosophie bei Platon, Descartes, Wolff und Lichtenberg,* Stuttgart: Metzler.

Schlick, M. (1932 [1969]) "Gibt es ein materiales Apriori?" in *Wissenschaftlicher Jahresbericht der Philosophischen Gesellschaft an der Universität zu Wien — Ortsgruppe Wien der Kant-Gesellschaft für das Vereinsjahr 1931/1932,* Wien: Verlag der Philosophischen Gesellschaft. Quotations are from M. Schlick, Gesammelte Aufsätze 1926-1936, Hildesheim: Olms 1969, 19-30.

Schöne, A. (1982) *Aufklärung aus dem Geist der Experimentalphysik: Lichtenbergsche Konjunktive,* Munich: C. H. Beck.

Shi-Hyong, K. (2008) *Bacon und Kant: Ein Erkenntnistheoretischer Vergleich zwischen dem `Novum Organum' und der `Kritik der reinen Vernunft',* Frankfurt: de Gruyter.

Stuart, M. T. (2017) "Imagination: A sine qua non of science," *Croatian Journal of Philosophy.*

Stump, D. J. (2011) "Arthur Pap's functional theory of the a priori," *HOPOS: The Journal of the International Society for the History of Philosophy of Science* 1: 273-289.

Tanona, S. (2010) "Theory, coordination, and empirical meaning in modern physics," in *Discourse on a New Method: Reinvigorating the Marriage of History and Philosophy of Science,* edited by M. Domski and M. Dickson, Chicago: Open Court.

Westphal, K. R. (2003) "Epistemic reflections and cognitive reference in Kant's transcendental response to skepticism," *Kant-Studien* 94: 135-171.

Westphal, K. R. (2004) *Kant's Transcendental Proof of Realism,* Cambridge: Cambridge University Press.

Wilson, A. D. (1998 [2014]) *"Introduction"* to the English Translation of Ørsted 1811 [1920], XV-XL.

Witt-Hansen, J. (1976) "H. C. Ørsted, Immanuel Kant, and the thought experiment," *Danish Yearbook of Philosophy* 13: 48-65.