DOI: https://doi.org/10.56550/d.4.1.1		Original research article
Received on: December 28, 2024	Accepted on: December, 11 2025	Published on: March, 24, 2026

Marco Buzzoni
University of Macerata
marco.buzzoni@unimc.it

KANT'S EXPERIMENTS OF REASON AND THOUGHT EXPERIMENTS IN PHILOSOPHY

Abstract

Kant did not distinguish between thought experiments (TEs) and real world experiments, nor did he use the German term "Gedankenexperiment". But many aspects of Kant's thought are related to the concept of TE, and especially "the experiments of reason" (*Experimente der Vernunft*), which are the distinguishing feature of transcendental arguments in philosophy as compared with the mathematical and empirical sciences. The first part of this paper reconstructs the Kantian concept of "experiments of reason". In the second part, a functional account of Kant's a priori is defended and contrasted with Ørsted's first Kantian interpretation of TEs. The entirely functional view of the a priori leads to an account of TEs in philosophy which is characterized by the following four features: 1) transcendental arguments, interpreted as philosophical TEs, are exemplifications of counterfactual reasoning that are expressions of the unlimited criticism typical of philosophy; 2) as far as their content is concerned, they depend entirely on information from 'outside', i.e. from common life experience and the empirical sciences, and it is this content that, albeit with respect to the conditions of possibility, grounds the criterion of their internal consistency; 3) apart from the reversal of the direction of enquiry, there is no difference in principle between the special methods of reasoning adopted in scientific and philosophical TEs (any such difference could only be based on a material conception of the a priori); 4) both philosophy and philosophical TEs have in the last analysis only a critical or negative task, that of denouncing all attempts to reduce the human person to categories or concepts derived, explicitly or implicitly, from empirical reality (whether natural or cultural).

Key words: Kant's Experiments of Reason; Transcendental Arguments; Thought Experiments in Philosophy; Functional Vs. Material A priori; Philosophy and Science; Thought Experiment and Human Person

KANTS EXPERIMENTE DER VERNUNFT UND GEDANKENEXPERIMENTE IN DER PHILOSOPHIE

Zusammenfassung

Kant hat weder zwischen Gedankenexperimenten und Realexperimenten unterschieden, noch hat er den deutschen Begriff „Gedankenexperiment" verwendet. Aber viele Aspekte von Kants Denken, insbesondere die „Experimente der Vernunft", sind mit dem Begriff des Gedankenexperiments verbunden. Die Experimente der Vernunft, die der erste Teil dieses Beitrags rekonstruiert, sind das Unterscheidungsmerkmal der Transzendentalphilosophie gegenüber der Mathematik und der empirischen Erkenntnis. Im zweiten Teil wird eine ohnehin funktionale Sichtweise von Kants Apriori verteidigt und mit Ørsteds erster kantischer Interpretation des Gedankenexperiments kontrastiert. Die funktionale Sicht des Apriori führt zu einer Darstellung der Gedankenexperiments in der Philosophie, das durch die folgenden vier Merkmale gekennzeichnet ist: 1) Philosophische Gedankenexperimente sind Beispiele kontrafaktischen Denkens, die Ausdruck der die Philosophie kennzeichnenden unbegrenzten Kritik sind; 2) was ihren Inhalt betrifft, so hängen sie vollständig von Informationen ab, die von „außen" kommen, d.h. von der allgemeinen Lebenserfahrung und den empirischen Wissenschaften, und dieser Inhalt begründet, wenn auch hinsichtlich der Bedingungen seiner Möglichkeit, das Kriterium ihrer inneren Konsistenz; 3) abgesehen von der Umkehrung der Untersuchungsrichtung gibt es keinen prinzipiellen Unterschied zwischen den speziellen Argumentationsmethoden, die in wissenschaftlichen und philosophischen Gedankenexperimenten angewandt werden (ein solcher Unterschied könnte nur auf einer materiellen Konzeption des Apriori beruhen); 4) sowohl die Philosophie als auch die philosophischen Gedankenexperimente haben letztlich nur eine kritische oder negative Aufgabe, nämlich alle Versuche zu widerlegen, welche die menschliche Person auf Kategorien oder Begriffe reduzieren, die aus der empirischen Realität (sei sie natürlich oder kulturell) implizit oder explizit abgeleitet sind.

Schlüsselwörter: Kants Experimente der Vernunft; Transzendentale Argumente; Gedankenexperimente in der Philosophie; funktionales vs. materielles Apriori; Philosophie und Wissenschaft; Philosophie und Wissenschaft; Gedankenexperiment und menschliche Person

Introduction

Kant did not distinguish between thought experiments (TEs) and real world experiments, nor did he use the German word "Gedankenexperiment". But many aspects of Kant's philosophy are related to the concept

of TE, and especially "the experiments of reason" (Experimente der Vernunft), which are the distinguishing feature of transcendental arguments in philosophy as compared with the mathematical and empirical sciences. In Section 2, the paper reconstructs the Kantian concept of "experiments of reason" as the method peculiar to transcendental philosophy. Section 3 sketches Hans-Christian Ørsted's first Kantian account of TEs, which overemphasized an aspect of Kant's a priori which, starting with Mach, will be sharply criticized by the philosophy of science of the twentieth century. This is probably the main reason why Ørsted's point of view had no important influence on the historical development of the concept of TE in the philosophy of science up to very recent times. Against this, I shall defend an interpretation of the a priori as purely functional, retaining the necessary and universal character of Kant's a priori and assigning no material content at all to it because, to put it now very simply and in a way that would be congenial to Kant, content can be given only by experience.

This is not the only way to pursue the project of a Kantian or Quasi-Kantian view of TEs. Others followed different paths. For example, Yiftach Fehige rejected the universality and necessity of Kant's a priori but retained the idea that the a priori, as a constitutive element of experience, is endowed with material content that may be made explicit by TEs. A very different example is the more recent Kantian account by Eleanor Helms, who maintains that in some cases TEs, even when they concern that which lies outside all possible phenomenal experience, can provide intuitive content (cf. Helms 2022). Both interpretations move in the opposite direction to the one I suggest here, which takes the Kantian rejection of an intellectual intuition in philosophy to its ultimate consequences, eliminating from the a priori all material content, no matter whether this latter is understood as contingent and relative (Fehige 2012 and 2013) or, following the opposite path of Ørsted's (and Husserl's), as something that can be known without recourse to experience.

In the Section 4 of this paper, a view of philosophical TEs is developed starting from Kant's notion of the "experiments of the reason", interpreted on the basis of a purely functional notion of the a priori. According to this view, TEs in philosophy (or transcendental arguments) can be characterized by four connected features: 1) they are exemplifications of counterfactual reasoning that express the same unlimited criticism of philosophy; 2) as far as their content is concerned, they depend entirely on information from 'outside', i.e. from common life experience and the empirical sciences, and

the criterion of their internal consistency is grounded in this content, albeit with respect to its conditions of possibility; 3) apart from the reversal of the direction of enquiry, there is no difference in principle between the methods of reasoning adopted in scientific and philosophical TEs or transcendental arguments (any such difference could only be based on a material conception of the a priori); 4) philosophical arguments or TEs have in the last analysis only a critical or negative task, that of denouncing all attempts to reduce the human person to categories or concepts derived from empirical reality (whether natural or cultural).

1. Kant's "experiments of reason"

Kant did not distinguish between thought experiment (TE) and real world experiment, nor did he use the German word "Gedankenexperiment", but he is nonetheless of great importance for the history and theory of TE. First of all, we find in his writings many concepts connected to the current debate about TEs in philosophy, such as "nur im Gedanken haben", "existi[e]ren nur im Gedanken" or the distinction between *Denken* (thinking) and *Erkennen* ("knowing" or "cognizing"). But still more important for our purposes is the fact that these expressions or concepts are, in turn, related to that of "experiment of reason" (*Experiment der Vernunft*), which deals directly and explicitly with the status of transcendental reflection as distinct from (but also related to) the natural sciences and mathematics.

The expressions "existi[e]ren nur im Gedanken" and "nur im Gedanken haben" recur in passages which bear upon the ability of the mind to methodically abstract from the matter of representations, in order to isolate the formal and a priori element of them. For example, in the *Metaphysische Anfangsgründe der Naturwissenschaft*, Kant notes that the parts of a phenomenon "exist only in thought", namely in the act of dividing a phenomenon:

> „Therefore, one can only say of appearances, whose division proceeds to infinity, that there are just so many parts in the appearance as we may provide, that is, so far as we may divide. For the parts, as belonging to the existence of an appearance, *exist only in thought*, namely, in the division itself. Now, the division does of course proceed to infinity, but it is still never given as infinite. Thus it does not follow, from the fact that its division proceeds to infinity, that the divisible contains an infinite aggregate of parts *in itself,* and outside of our representation." (Kant 1900 ff., AA, IV, 506-507, Engl. transl. from Kant 1786 [2004], p. 44; italics is mine.)

In the same work it is also said that the concept of absolute space arises from our ability to abstract from the matter of a phenomenon, since, like that of part, it can be given only "in thought" without its corresponding object (cf. Kant 1900 ff., AA, IV, 481-482; Engl. transl. from Kant 1786[2004], p. 14).

In general, the ability to abstract from matter and to have (only) in thought the pure formal aspects of a phenomenon is one of the many complex aspects of the distinction between "thinking" (as a translation of "Denken") and "knowing" or "cognize" (as translations of "Erkennen" by Smith and, respectively, Guyer and Wood). One of the most important passages in which Kant draws this distinction is at the beginning of § 22 of the second edition of the *Critique of pure reason* (cf. respectively Kant 1787[1929], B 146, pp. 161-162, and Kant 1787[1998], B 146, p. 254), but it is well known that the distinction plays a key role in Kant's rejection of the ontological argument. According to *Der einzig mögliche Beweisgrund zu einer Demonstration des Daseins Gottes*, the argument would be based on the confusion between "to exist" and "to be conceived of as existing," between actual existence and "existence only in thought" (*gedachte[r] Existenz*): real existence can be clearly distinguished from merely imagined existence, and the relationship between these two types of existence is illustrated by the famous example of the 100 thalers (Kant 1900 ff. AA, II, 156, line 14).

All these distinctions are intimately connected with the concept of "experiment of reason" (*Experiment der Vernunft*), which is introduced in the *Preface* to the second edition of the *Critique or pure reason.* This concept, with very few exceptions, has received little attention in the literature both on Kant's philosophy in general and on TEs in particular (cf., for Kant in general, Kalin 1972; Westphal 2003 and 2004; Kühne 2005; Shi-Hyong 2008; for TEs, see Buzzoni 2011, 2018 and 2019; Fehige 2012; Fulkerson-Smith 2013; Fehige and Stuart 2014, Schmid 2020). Both cases are difficult to explain. The first case, both because of the place in which the concept was formulated (probably one of the most often cited places in the history of philosophy) and because it deals with the distinction in principle between philosophy (not only as a critique of pure reason, but also as a critique of practical and teleological reason) and scientific-empirical or mathematical reasoning; the second case, because the expression in itself should have suggested an important connection with the current debate on TEs in philosophy. A brief analysis of the concept of "experiments of reason" easily confirms both of these points.

According to Kant, the transcendental critique of reason (which analyses the a priori formal conditions of the possibility of experience) only avoids the charge of dogmatism or arbitrariness if there is a criterion distinguishing true and false arguments. This criterion is given by the "method" of the "experiments of reason".

Let us first look at Kant's most important passages about the experiments of pure reason, first of all two long footnotes explaining by what method "critical metaphysics" can set itself on the sure path of a science. Here is the first one:

> "This method, imitated from the method of those who study nature, thus consists in this: to seek the elements of pure reason in that **which admits of being confirmed or refuted through an experiment.** Now the propositions of pure reason, especially when they venture beyond all boundaries of possible experience, admit of no test by experiment with their **objects [***Objekte***]** (as in natural science): thus to experiment will be feasible only with **concepts** and **principles** that we assume *a priori* by arranging the latter so that the same objects can be considered from two different sides, **on the one side** as objects of the senses and the understanding for experience, and **on the other side** as objects that are merely thought at most for isolated reason striving beyond the bounds of experience. If we now find that there is agreement with the principle [*Princip*] of pure reason when things are considered from this twofold standpoint, but that an unavoidable conflict of reason with itself arises with a single standpoint, then the experiment decides for the correctness of that distinction." (KrV, B xviii—xix; Engl. transl. from Kant 1787[1998], p. 111; cf. Kant 1900 ff., AA, III, 13 Anm.)

The second note better clarifies the analogy and difference between Kant's "critical metaphysics" on the one hand and the experimental sciences or mathematics on the other:

> "This experiment of pure reason has much in common with what the **chemists** sometimes call the experiment of **reduction**, or more generally the **synthetic procedure**. The **analysis of the metaphysician** separated pure a priori knowledge into two very heterogeneous elements, namely those of the things as appearances and the things in themselves. The **dialectic** once again combines them, in **unison** with the necessary rational idea of the **unconditioned**, and finds that the unison will never come about except through that distinction, which is therefore the true one." (KrV, B xxi; Engl. transl. from Kant 1787[1998], p. 112; cf. Kant 1900 ff., AA, III, 14, Anm.)

This difference is taken up later in the *Critique*, in the "Doctrine of method" (and more specifically in the section devoted to "The discipline

of pure reason in dogmatic use"), as the difference between, on the one hand, the "acroamatic (discursive) proofs" of philosophy and, on the other hand, contrary to Ørsted's later opinion, mathematical "demonstrations" (*Demonstrationen*):

> „Philosophical cognition [...] must do without this advantage, since it must always consider the universal *in abstracto* (through concepts), while mathematics can assess the universal *in concreto* (in the individual intuition) and yet through pure *a priori* intuition, where every false step becomes visible. Since they can only be conducted by means of mere words (the object in thought), I would therefore prefer to call the former **acroamatic** (discursive) proofs rather than **demonstrations**, which, as the expression already indicates, proceed through the intuition of the object." (KrV, B 762-763; Engl. transl. from Kant 1787[1998], p. 112; cf. Kant 1900 ff., AA, III, 481-482.)

The quoted passages make clear both the similarity and the difference between the experiments of the natural sciences and mathematics, on the one hand, and those proper to transcendental philosophy, on the other. In both cases it is a kind of experimentation, but while physical experimentation takes place on concrete objects (which can be known or cognized: *erkannt*), the experiments of reason (as belonging to philosophy) are what today we would call TEs, since their objects are only "thought". Since, for the testing of its propositions, no real world experiment can be carried out with the objects of pure reason (which, on the other hand, is possible with the objects of the natural sciences), we need a different criterion for assessing the ability of human reason to abstract from the material of knowing, i.e., to have the formal moments only "in thought" (*in Gedanken*) and judge them correctly. For this purpose, the method used in the *Critique of pure reason* consists in considering objects from two different points of view: on the one hand, as known objects or appearances, and, on the other, as objects "merely thought" (as things in themselves or noumena, i.e., as limiting concepts). The "experiment of pure reason" decides in favour of the correctness of this distinction, if, using it, we do not fall into contradiction, while, on the contrary, if we unilaterally adopt only one of these points of view, reason inevitably finds itself entangled in a "self-conflict."

The essential connection of the "experiments of pure reason" with the concept of antinomies is evident (and in this sense Kant designates the method of the experiments of pure reason as a "skeptical method"), and it is therefore quite natural that the following passage concerning the "experiments of (pure) reason" is found precisely in the course of Kant's discussion

of the antinomies. The passage contrasts transcendental philosophy as such with both physics (here called, as often in Kant, "experimental philosophy": *Experimentalphilosophie*) (KrV, B, 452-453; Engl. transl. from Kant 1787[1998], p. 469; cf. Kant 1900 ff., AA, III, 292-293) and mathematics:

> "This skeptical method [...] is essentially suited only to transcendental philosophy, and can in any case be dispensed with in every other field of investigation, but not in this one. In mathematics its use would be absurd, because nowhere in mathematics do false assertions disguise themselves and make themselves invisible; for mathematical proofs always have to proceed along the lines of pure intuition, and indeed always through a self-evident synthesis. In experimental philosophy a doubt postponing judgment can be useful, but at least there is no possible misunderstanding that cannot be easily removed, and the ultimate means for deciding the controversy must at last lie in experience, whether it is found early or late. [...] On the contrary, the transcendental assertions that presume to extend their insight beyond the field of all possible experience are neither in the case where their synthesis could be given in an *a priori* intuition, nor are they so constituted that a misunderstanding could be exposed by means of any experience. Transcendental reason thus permits no touchstone other than its own attempt to bring internal unification to its assertions, and this requires a free and unhindered contest of these assertions among themselves, which we will now initiate." (KrV, B 452-453; Engl. transl. from Kant 1787[1998], p. 469; cf. Kant 1900 ff., AA, III, 292-293)

From this point of view, the entire *Critique of pure reason* is *strictu sensu* an "experiment of reason", that is, in the last analysis only one philosophical TE whose truth is guaranteed by the fact that – as the famous Kantian thesis goes – human understanding gets lost in antinomies when it ventures beyond the limits of possible experience and attempts to deal with things in themselves.

In this sense, Kalin's thesis that Kant's transcendental arguments "can be characterized [...] as Gedankenexperimente" (Kalin 1972, p. 322) is essentially right, but, first, as we have just seen, it should be applied to the entire *Critique of pure reason* and, second, as we shall see now, should be extended to Kant's *Critique of practical reason* and *Critique of judgment*, which both contain an antinomy.

First of all, the experiments of reason are instantiated in the first formulation of the categorical imperative: act only according to the maxim that one can will, without contradiction, that it should become a universal law of nature (cf. e.g. GMS, AA, IV, 421, 07-08, Engl. Transl. by Lewis White

Beck, Kant 1785[1997], p. 38; for the formulation of the categorical imperative as a TE, see Illies 2007, p. 313).

Secondly, and most importantly, in a passage of the *Critique of practical reason* Kant talks explicitly about an "experiment" of practical reason (and he draws the same comparison with a chemist as that already found in the *Critique of pure reason*). In the Analytic of pure practical reason the philosopher, Kant writes, has to struggle with greater difficulties than the geometrician because he does not have any intuition of the noumenon, but he nevertheless can proceed "almost like a chemist":

> "almost like a chemist" (*beinahe wie der Chemist*) he can at any time set up an experiment with every human practical reason in order to distinguish the moral (pure) determining ground from the empirical, namely, by adding the moral law (as a determining ground) to the empirically affected will (e.g., that of someone who would gladly lie because he can gain something by it). When an analyst adds alkali to a solution of calcareous earth in hydrochloric acid, the acid at once releases the lime and unites with the alkali, and the lime is precipitated. In just the same way, if a man who is otherwise honest (or who just this once puts himself only in thought in the place of an honest man) is confronted with the moral law in which he cognizes the worthlessness of a liar, his practical reason (in its judgment of what he ought to do) at once abandons the advantage, unites with what maintains in him respect for his own person (truthfulness), and the advantage, after it has been separated and washed from every particle of reason (which is altogether on the side of duty), is weighed by everyone, so that it can enter into combination with reason in other cases, only not where it could be opposed to the moral law, which reason never abandons but unites with most intimately." (*Kritik der praktischen Vernunft*, AA, V, 92-93, lines 33-37 and 1-10; Engl. Transl. by M. Gregor and A. Reath, Kant 1788[2015], pp. 75-76. In this case, we might perhaps also entitled to speak of 'experiments of pure practical reason').

Finally, Kant uses the expression "Experiment der reinen Vernunft" in the essay *Zum ewigen Frieden*. 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 realised in any particular case? By means of "an experiment of pure reason", which provides the requested criterion a priori. We have to judge whether the "form" of the law can or cannot be consistently combined with the principles of a moral agent. If not, we immediately recognize the falsity, that is, the illegality of a legal claim:

> „Any legal claim must be capable of publicity. The capability of publicity can therefore, since one can quite easily judge whether it obtains in a given case, that is, whether or not it is consistent with the basic principles of the agent, provide an easily applicable criterion that is found *a priori* in reason. If it is not consistent with the agent's principles one can recognize through an experiment of pure reason, as it were, the falseness (opposition to the law) of any given claim (*praetensio iuris*)." (*Zum ewigen Frieden. Ein philosophischer Entwurf*, AA, VIII 381, 4-18; Engl. Transl. by David L. Colclasure, Kant 1975[2006], p. 104.)

Kant, in short, on the one hand emphasizes the analogy of transcendental (including ethical-practical) with chemical arguments, but it is peculiar trait of the experiments of reason as transcendental arguments that they separate counterfactually a material and a formal part and have to resolve any contradiction of reason with itself.

2. Ørsted's Kantian Account of TE.

After the previous essential analysis of the most important Kantian passages devoted to the "experiments of reason", in the second part of my paper I shall make use of the indications contained therein to arrive at a conception of philosophical TEs as far as possible in accordance with what I regard as the spirit of Kant's philosophy. For this purpose, I shall accept Kant's claim that philosophy or critical metaphysics, unlike the natural sciences, cannot establish the truth of a statement simply by experimenting. Moreover, I shall accept Kant's assertion that in order for philosophical (transcendental) arguments not to be arbitrary, there must be a criterion by which true arguments can be distinguished from false ones.

However, in order to delineate a consistent conception of TEs in philosophy inspired by Kant's critical philosophy, it is necessary, before proceeding further, to clear up an ambiguity concerning Kant's conception of the a priori.

This ambiguity – which I can explore here very briefly (for further discussion of this point, see Buzzoni 2011, 2013, and 2018) – is about two distinct and ultimately contradictory interpretations of the a priori. On the one hand, Kant's account of the a priori is a merely formal or functional one. He explicitly says that the *a priori* determines only the formal conditions of experience and is devoid of material content: categories, if not applied to sense contents through schemata, are "only functions of the understanding for concepts" that "do not represent any object" (KrV, B 187 and KrV, B

305, Engl. transl. from Kant 1787[1998], p. 277; cf. Kant 1900 ff., AA, III, 139). A merely transcendental use of the categories has no determinate object, not even one that is determinable in its mere form. If separated from all sensibility, pure categories have no use, either empirical or transcendental: they "are not supposed to have empirical use, and cannot have transcendental use, they do not have any use at all if they are separated from all sensibility, i.e., they cannot be applied to any supposed object at all" (KrV, B 305, Engl. transl. from Kant 1787[1998], p. 390; cf. Kant 1900 ff., AA, III, 208). The same considerations are valid for the "I think", which, as the supreme condition of the possibility of experience, cannot have a content of its own:

> "**I** do not cognize [*erkenne*] any objects merely by the fact that **I** think [*denke*], but rather **I** can cognize [*erkennen*] any object only by determining a given intuition with regard to the unity of consciousness, in which all thinking consists. Thus **I** cognize [*erkenne*] myself not by being conscious of myself as thinking [*als denkend*], but only if **I** am conscious to myself of the intuition of myself as determined in regard to the function of thought [*Funktion des Denkens*]. All *modi* [...] of self-consciousness in thinking [*im Denken*] are therefore not yet themselves concepts of the understanding of objects [...] (categories), but mere functions, which provide thought [*Denken*] with no object at all, and hence also do not present my self as an object to be cognized [*to erkennen geben*]." (KrV, B, 406-407, Engl. transl. from Kant 1787[1998], p. 445)

On the other hand, however, it is to some extent undeniable that Kant ascribes some content to the a priori, for example insofar as he believes that transcendental philosophy can provide a genuine metaphysical foundation for Newton's physics and that a part of Newton's physics provides us with universal and necessary knowledge that needs no further substantive modification (cf. the "pure part" of Newton's Principia according to Kant's MAN).[1] Thus, in a well-known passage of *Metaphysische Anfangsgründe der Naturwissenschaft*, Kant demonstrates a priori that only attraction and repulsion can be considered as moving forces of matter:

> "Only these two moving forces of matter can be thought. For all motion that one matter can impress on another, since in this regard each of them is considered only as a point, must always be viewed as imparted in the straight line between the two points. But in this straight line there are only two possible

[1] On this point, cf. Friedman 1992, p. 4. That Kant considered Newtonian physics an essentially ultimate cognitive acquisition is a thesis that many philosophers of science have also insisted on. See, to cite one of the most clearest examples, Popper 1963, ch. 2, § 10-11 and ch. 7.

> motions: the one through which the two points *remove* themselves from one another, the second through which they *approach* one another. But the force causing the first motion is called *repulsive force*, whereas the second is called *attractive force*. Therefore, only these two kinds of forces can be thought, as forces to which all moving forces in material nature must be reduced." (MAN, in Kant 1900 ff., AA, IV, 498-499, Engl. transl. from Kant 1786[2004], pp. 35-36)

As rightly pointed out by Kühne 2005, in this passage Kant presents "some substantive statements of physics as a priori truths"[2]. To put this in our terms, Kant's argument betrays a content-based way of conceiving the a priori. In fact, it is a typical example of pre-Kantian metaphysical speculation on nature, which claims to be reliable without being controlled by experience.

Both of these lines of thought (functional and material) in conceiving the a priori are well exemplified in Kant's conception of scientific theories. The problem is at the heart of the transcendental deduction and concerns the relationship between the a priori laws that the categories prescribe a priori to the manifold of experience and the particular physical laws. On the one hand, Kant is fully aware that transcendental philosophy cannot take the place of physics:

> "The pure faculty of understanding does not suffice, however, to prescribe to the appearances through mere categories *a priori* laws beyond those on which rests a **nature in general**, as lawfulness of appearances in space and time. Particular laws, because they concern empirically determined appearances, **cannot** be **completely derived** from the categories, although they all stand under them. Experience must be added in order to come to know particular laws **at all**; but about experience in general, and about what can be cognized as an object of experience, only those *a priori* laws offer instruction. (KrV, B 165, Engl. transl. from Kant 1787[1998], pp. 263-264; cf. Kant 1900 ff., AA, III, 127)

But on the other hand all phenomena, that is, everything that can come to empirical consciousness, are subject, as to their unification, to the

[2] Cf. Kühne 2005, pp. 104-105. On this point see also Friedmann 1992, e.g. pp. 174-177, and Tanona 2000, especially pp. 436-440. This latter writes: "For a science to obtain [...] certainty, it must have a special metaphysics that provides a foundation for it. The foundation for physics in particular must be very broad and must provide a metaphysics for the concept of matter in general; but despite this broadness it is still a special metaphysics that goes beyond the realm of the Critique. The Metaphysical Foundations treats the empirical concept of matter in general and analyzes its content to provide an a priori foundation for the physics of matter." (Tanona 2000, 437)

categories, on which the laws of nature depend, and thus are in line accessible to what we have seen Kant define as "Experimente der Vernunft," proper to the transcendental method:

> "Now the possibility of cognizing [erkennen] *a priori* **through categories** whatever objects **may come before our senses**, not as far as the form of their intuition but rather as far as the laws of their combination are concerned, thus the possibility of it as were prescribing the law to nature and even making the latter possible, is to be explained. For if the categories did not serve in this way, it would not become clear why everything that may ever come before our senses must stand under the laws that arise *a priori* from the understanding alone." (KrV, B 159-160, Engl. transl. from Kant 1787[1998], p. 261; cf. Kant 1900 ff., AA, III, 124)

More in general, Kant's claim that the "I think" is the legislator of nature is in itself ambiguous. It does not have a precise meaning unless we are given an (a priori) specification of certain general laws that form the content of its "legislation." Kant's distinction between some very general laws, deducible a priori, and empirical laws, derivable only by empirical means and by experiment, can only be drawn if the "I think" possesses its own logical-transcendental content, which can be imposed on that of the senses and could be specified a priori. But if one accepts a purely functional conception of the a priori, the Kantian claim that the "I think" imposes its own laws on nature can be accepted only in the very general sense that all representation and law must be synthesized a priori by it. In other words, if understood in a purely functional way, the "I think" cannot impose anything on phenomena except the 'fact' of being something to us. This being something to us is indeed irreducible to empirical phenomena, but does not add to them any concrete determination, any particular content. In fact, if Kant's transcendental deduction is to remain distinct from the Hegelian deduction, it cannot, *to any extent*, bring particular phenomena back to determinate laws, but must confine itself to justify the a priori possibility of knowing reality so far as it is actually given in determinate conditions, apart from our reason or will. Insofar as it brings particular phenomena back to determinate laws, to that extent Kant's transcendental deduction falls within (and anticipates the) Hegelian concept of deduction.

As a special case of this general difficulty, I mention that the simple fact that the categories are twelve - i.e. a manifold, a finite and definite number – implies that they can be distinguished on the basis of a different content. If they were pure functions, they could not be distinguished (and then

counted) because there would be no content to be distinguished. Kant thus neglected what he himself had noted in the Logic: if the multiplicity of the a priori forms of understanding (and the same naturally applies to those of intuition) is assumed, it is necessary to accept their "kind and number" as a mere fact, in contradiction with the purely formal-transcendental truth-value of the arguments that argue for the a priori forms. Every division - Kant had written - that "has *more than two* members", i.e. every "polytomy", is "empirical" and cannot be "from principles a *priori*": [...] Polytomy requires intuition; either *a priori*, as in mathematics (e.g. in the division of conic sections), or empirical intuition, as in description of nature." (*Logic*, AA, IX, § 113; Engl. Transl. by J.M. Young, Kant 1800[1992], p. 637)

Now, Hans Christian Ørsted's account of TEs was largely based on the residue of the material aspect of the a priori that persists in Kant. Ørsted introduced the expression "TE" 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. Strictly speaking, the first 'Kantian' inspired theory of TE was not mine (cf. Brown 2011, p. 202), but that of Hans Christian Ørsted, who however developed a theory of TEs based on a material, not functional, conception of the Kantian a priori, that is, in a direction opposite to that I have pursued. Because Ørsted's concept of "Gedankenexperiment" was strongly conditioned by a material interpretation of Kant's a priori, it tended to blur the distinction in principle between mathematics, physics, and philosophy, at least to the extent that they all constitute versions of one and the same practice of the a priori physics we have already encountered in Kant's *Metaphysische Anfangsgründe der Naturwissenschaft*.

According to Christian Ørsted, the method of "Gedankenexperiment" is a construction of concepts in intuition which, widely used in mathematics, should play an important role in the development of empirical theories:

> „If we let a point move in our representation in order to describe a line, or let a line move around its own end point in order to describe a circle with the other, what is this other than a thought experiment? Differential and integral calculus consists entirely of such thought experiments and considerations. [...] Here we see every truth in its genesis. The ground of its existence and our certainty therefore coincide; so that when it is represented in this way, it is at the same time already proved. Kant has given us the most beautiful examples of

> this mode of representation in his Metaphysical foundations of natural science, but without carrying out the consideration developed here."[3]

In light of these (and other passages), Kühne's thesis that Ørsted's *Ideen zu einer neuen Architektonik der Naturmetaphysik* sought to extend the *Metaphysische Anfangsgründe* to both the external behaviour of motion and the physical and chemical properties of substances (cf. Kühne 2005, p. 116), seems essentially correct. With his remarks on the method of the TE, Ørsted aimed at "a further development of Kant's transcendental foundation program", through which it was possible to transform important sections of the theory of nature into mathematics (cf. Kühne 2005, p. 135).

From the point of view defended here, this is the natural consequence of the tendency we have shown in Ørsted towards a material conception of the a priori. Our claim can be supplemented and even counter-checked (also providing further evidence for the substantive correctness of Kühne's interpretation) by the fact that the aforementioned Kantian ambiguity concerning the distinction between a priori and particular natural laws is cleared up by Ørsted by dispensing with the contribution of experience, i.e. by implicitly assuming a material a priori, which grounds the necessity and universality (i.e., in Kant's lexicon, the a priori character) of the laws of nature:

> "*Kant* does not want to deduce the laws of external intuition absolutely *a priori*, but states that in every part of the metaphysics of nature another feature given by experience must be added. This statement, however, seems to contradict the proposition established and proved by *Kant* himself, *that the laws of nature must be general and necessary*, and therefore cannot be deduced from experience. It is also evident to all that the universality and necessity of the laws of nature cannot consist in a deduction from experience." (Ørsted 1802, p. 10; cf. also Ørsted 1822, p. 485, where Ørsted says that by the method of TE "our science ought to be transformed into a mathematics of nature that would far surpass the previous one in both form and content.")

Ørsted's account of TE is a one-sided, though *prima facie* coherent, development of one aspect of the a priori in Kant: TEs become a way of justifying a new mathematical discipline, capable of proving propositions endowed with physical-chemical meaning (which is tantamount to an a priori physical or chemical science). In this way, Ørsted went against the trend prevalent in the philosophy of science since the late 1800s, whose

[3] Ørsted 1822, p. 482-483. As for the prehistory of the concept of "thought experiment," one should at least mention Lichtenberg (cf. Lichtenberg 1968[1994] and 1971[1994]. On Lichtenberg, see especially Fehige and Stuart 2014.

main representatives rejected any material interpretation of the Kantian a priori that is connected with a claim of universal and necessary truth-value. Among these philosophers of science – besides Reichenbach and the logical empiricists or, even earlier, Ernst Mach and Jules Poincaré – we can name, for example, Karl Popper, Konrad Lorenz, Clarence Irving Lewis, Arthur Pap, Israel Scheffler, Stephan Körner, John Dewey and Thomas Kuhn (Lipton 2003: Kuhn is «Kant on wheels»).

The well-documented fact that the entire history of the philosophy of science is strongly characterized by the rejection of the Kantian a priori provides the most plausible explanation why Ørsted's interpretation of TE – unlike that of Ernst Mach – has been neglected until recent years, while Mach's conception found an immediate echo in the philosophy of science.[4] In other words, from the point of view of our previous consider-

[4] Recently, however, Helms 2022 has argued that Ørsted view had an important influence on the development of the concept of TE because of his influence on Mach, given that Mach could not have been unaware of Ørsted's writings and that there are striking similarities between Ørsted's and Mach's view about TE. Now, because the first claim is probably true, while the second is arguably false, all we are entitled to suppose is that Mach assumed from Ørsted the word, but not the corresponding concept, of "Gedankenexperiment". The second claim is arguably false because the difference between Mach's and Ørsted's view is much more philosophically important than the four similarities Helms notes. According to her, for both Ørsted and Mach: 1) the idea of scientific experiment is based on the experimental method of variation; (2) scientific thought is an expression of human development towards autonomy and freedom; (3) TE is able to solve the problem of new knowledge in geometry; (4) TEs are important for didactic purposes. First of all, these similarities are not specific enough to bear on an answer to the question of the influence of Ørsted on Mach. For example, that experiments are based on the method of variation is certainly not enough to show Ørsted's influence on Mach, since already Bacon, Galileo and Mill, to whom Mach made repeated references in his works, had this idea of experiment. For example on pp. 4 and 388-389 of the *Principien der Wärmelehre* (1900 edition) Mach quotes Mill's methods and refers to the principle of the thermometer as based on the variation of the volume of the liquid depending on the temperature. Secondly, and most importantly, even granting some relevance to these similarities, they would not exclude the essential difference of Mach's and Ørsted's philosophical attitude towards Kant's a priori. While Ørsted's is the first relevant Kantian conception of TEs, Mach's soon became the exemplary model for all subsequent empiricist interpretations of TEs. This did not happen by chance. Mach's explicit rejection of Kant's philosophy is well known (cf. e.g. Mach 1903[1914], pp. 299-300, and Mach 1905[1976], Engl. Transl., p. 23 and 32 fn.). More specifically, Mach rejected of Kant both the thing in itself (cf. e.g. Mach 1903[1914], p. 292), and any possible a priori knowledge (including the category of cause or the nature of mathematical knowledge) whose value may be independent of experience and does not

ations, the one-sided development of the Kantian concept of the a priori in a material sense largely explains why the recognition of Ørsted's conception of TEs came so late, so that he, unlike Mach, had almost no influence on the development of the concept of TE in the twentieth-century philosophy of science. This latter mostly rejected the Kantian a priori in Ørsted's material sense, either in the empiricist direction of Mach, logical empiricism and Popper, or in the conventionalist direction of French philosophy, of the relativistic philosophy of science of the 1960s and of the sociological turn. That there are determinate parts of our knowledge (such as physics and mathematics) that can be founded directly on synthetic a priori principles seemed to almost all the major exponents of the philosophy of science to be refuted by the history of science itself: relativistic physics, quantum physics, the rise of non-Euclidean geometries, all seemed to prove that there are no a priori principles endowed with contents immune from critical revision by experience or not depending upon conventions.

Moritz Schlick, in particular, explicitly criticized the phenomenological development of the Kantian a priori as a "material a priori," that is, an a priori that possesses contents immune from any critical revision. According to Schlick, the question "Is there a material a priori?" must be answered with a clear "No!": "The empiricism I advocate believes it knows 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 1969, p. 25). In the same vein, Hans Reichenbach specified two meanings of the notion of a priori knowledge in Kant, of which only the second, endowed with mere pragmatic value, was defensible: "Kant's concept of a priori has two different meanings. First, it means "necessarily true" (*apodiktisch gültig*) or "true for all time," and secondly, "constituting the concept of object"" (Reichenbach 1920[1965], p. 46, Engl. Transl. p. 48).

depend on empirical findings (cf. e.g. Mach 1905[1914], pp. 280-281, Engl. Transl., pp. 205-206), or is not reducible to an a priori of a biological kind very similar to that of Spencer and today's evolutionary epistemology (cf. e.g. Mach 1905[1914], p. 275, Engl. Transl., p. 206). Finally, unlike Ørsted's, Mach's concept of freedom is surely not Kantian: his conception of the ego is much closer to Hume and Mill than to Kant. He reduces the transcendental unity of the "I think" to a "unity of mental economy" ("denkökonomische Einheit") of sensations that change slowly over time (Mach 1903[1914], pp. 3, 7, 18-19). In sum, Ørsted's position is quite distinct from Mach's, because, unlike Mach, it is based on a Kantian notion of the a priori that, anticipating the following philosophy of science, Mach had explicitly rejected.

I think this rejection of the Kantian a priori was correct in its *pars destruens*, that is, insofar as it targeted the material aspect of the a priori on which Ørsted built his theory of TE. But this criticism was also used – in the spirit of Reichenbach's pragmatic a priori – as the *pars construens* of a different material view of the a priori. In this view, the material a priori is not universal and necessary, but particular and contingent. This meaning of the a priori has been revived and defended by many authors, among whom Michael Friedman deserves special mention (cf. e.g. Friedman 1992), not only as the sole legitimate source of belief, but also as an interpretation of the a priori that can still be legitimately regarded as Kantian. With regard to the understanding of TEs, it was mainly Yiftach Fehige who defended the importance of a relativized and contingent a priori. Rejecting the universality and necessity of Kant's a priori, he retains Reichenbach's (and Friedman's) idea of the a priori as a constitutive but theory-relative and contingent element of experience, endowed with a material content that can be made explicit by TEs (cf. Fehige 2012 and 2013).

This view, besides having certain serious theoretical shortcomings, does not seem to me to be Kantian either according the letter or the spirit of Kant's philosophy. I have argued for this elsewhere (cf. Buzzoni 2013) and, for reasons of space, will not dwell on it here. On the contrary, accepting the almost unanimous view of the tradition in the philosophy of science, in what follows I shall reject any material conception of the a priori, and shall instead move in the direction of a formal or, better, functional a priori, which is in my opinion the fundamental and most promising line of thought in Kant. More precisely, by taking the Kantian rejection of intellectual intuition to its ultimate and coherent consequences, I set out to develop a view that eliminates all content from the a priori, whether this content is understood as contingent and relative (Fehige) or whether one follows the opposite path (already defended by Ørsted, but later reaffirmed independently by Husserl), which attributes to human reason some capacity to intuit it (in this sense, see the already mentioned proposal by Helms 2022). Thus my position is, on the one hand, in agreement with the rejection of the material a priori proper to the twenty-philosophy of science, but, on the other hand, it departs from it in that it accepts, though only in a purely functional sense, the universality and necessity of the Kantian a priori.

3. Towards a Kantian Account of TEs in Philosophy

Now, in light of the considerations so far, I return to Kant's "experiments of reason", which I regard as an anticipation of TEs in philosophy that is worthy of being taken up and developed. I shall interpret both the Kantian conception of transcendental arguments and the nature of the "experiments in pure reason" from the viewpoint of the purely functional nature of the a priori.

Let us start first with Kant's conception of philosophy. In one of the most important definitions Kant offers of (transcendental) philosophy, he states that the latter consists of a system of concepts that "is occupied not so much with objects but rather with our mode of cognition of objects insofar as this is to be possible a priori." (KrV, B 25, Engl. transl. from Kant 1787[1998], p. 149; cf. Kant 1900 ff., AA, III, 43) In other words, philosophy reverses the direction or attitude adopted by scientific inquiry toward reality: it is not directed toward objects, but toward the conditions of possibility of the human reason's ability to stand in relation to reality.

What consequences follow for the relationship between philosophy and the empirical sciences (natural or human) from this way of conceiving philosophy? First of all it follows that philosophy is a type of inquiry without a priori limits *a parte objecti*, i.e., with respect to its possible objects. Whereas the empirical sciences shed light on particular aspects of reality and necessarily ignore others (if we investigate a philosophy book from the standpoint of Newtonian physics, we will only grasp properties of it such as mass or position in space and time, and ignore, for example, its importance in cultural history), philosophy cannot restrict its investigation in this way. For this reason, there is no physics of physics or physics of philosophy, while it makes perfect sense to speak of both a philosophy of physics and a philosophy of philosophy.

This unlimited character of philosophy *a parte objecti*, which results from the reversal of the direction of scientific inquiry, is the most important difference in principle between philosophy, on the one hand, and the empirical or formal-mathematical sciences, on the other. But this is, so to speak, only one side of the coin, the other being a limitation: if there is no limit to the possible objects of philosophy, strictly speaking, philosophy has no particular domain of objects of its own (a claim of the logical empiricists that is in a certain sense true, only not in the one-sided sense in which they upheld

it) and it must find its object outside itself, that is, in everyday life or in the natural, formal-mathematical or human sciences.

But what about the particular methods of science and philosophy? One might be inclined to think that the fact that philosophy reverses the direction of scientific inquiry entails important differences between the set of particular methods of philosophy and science. But the functional nature of the a priori, its being empty of material content, strictly implies that, apart from the direction of inquiry, any other differences between science and philosophy can arise only from differences in the object being investigated, not from particular characteristics that we should establish a priori. Briefly put, the fact that philosophy reverses the direction or attitude of scientific inquiry entails only differences which can ultimately be traced to whether or not the real world experiment is used or not (for more details on this point, see Buzzoni 2021).

However, this complete methodological naturalism must be accompanied by an anti-naturalist claim with respect to the irreducible status of philosophy as such. As we have already seen, according to Kant, transcendental reflection, through the experiments of pure reason, must succeed in maintaining a nexus of unity and distinction between phenomenon and noumenon, between the point of view of the "I think" and that of the "thing-in-itself", in short, between the (transcendental) subject and the (transcendental) object of knowing. Any full identification or separation between these distinct points of view leads to an antinomy, that is, to a pair of contradictions whose thesis and antithesis correspond to the subject's point of view and the object's point of view, respectively.

However, we have to criticize the particular way in which Kant understood this unity and distinction of points of view, which, as we have pointed out earlier, was influenced by the ambiguity between material and functional aspects of the a priori. Jacobi's best-known objection against Kant's critical philosophy provides a fine example of this: as he famously said, without the thing-in-itself it is impossible to enter the *Critique of pure reason*, but with the thing-in-itself it is impossible to stay within it (Jacobi 1815[1994], p. 304, Engl. Transl., p. 336).

This is precisely one of those counterfactual arguments that express a contradiction of reason with itself and that are, according to Kant, the distinctive trait of philosophical reasoning and transcendental arguments. This is true not only of Jacobi's critique, but also of the objections already raised (and many other objections that might be raised) against Kant's philosophy.

As already mentioned, for example, specifying the a priori and assuming it to be a multiplicity (of categories or, more generally, of the a priori forms of reason) is a thesis that must be rejected because it contains a contradiction: to retain the characters of universality and necessity of the a priori, one must abandon that of its multiplicity.[5]

This is also true of the more general criticism we raised against Kant's inconsistency of assuming both a material and a functional role of the a priori. And let us note by way of a counterproof that Jacobi's critique can be overcome by a purely functional conception of the a priori. The contradiction formulated by Jacobi can be removed by understanding the distinction between phenomena and things-in-themselves not as a contrast between different entities, but as a purely functional distinction indispensable for all empirical research, that between what is already known and what is not yet known.

Another interesting example is the objection, raised by some authors from a personalistic point of view, that Kant's a priori (and notion of "transcendental") is incompatible with the uniqueness and unrepeatability of the human person (cf. e.g. Stefanini 1959, p. 9).

In a sense, this objection is justified, but it can be avoided by distinguishing between different senses of the a priori. If the Kantian notion of a priori is understood as something possessing its own determinate content, it conflicts with the uniqueness (and freedom) of the person. In this case, universality and necessity of the a priori denies the uniqueness (and freedom) of personal beings because they are accompanied by the particular content of determinate structures or forms, which are superimposed upon the uniqueness of the human being.

However, from the point of view defended here, which denies any determinate content to the a priori, the objection is without any force. If one assumes the point of view of a purely functional a priori and maintains it more consistently than Kant, the irreducibility of the human person is merely functional, that is, it is never given once and for all, but must always be regained in the face of the never entirely predictable multiplicity of

[5] It should be pointed out that this argument not only does not exclude but implies multiplicity in a non-transcendental, empirical or positive sense, i.e. in the sense of a set of contingent and historically variable assumptions that we have to accept, reject or modify for the particular purposes of cognition. This sense, which was especially emphasized by the leading exponents of hermeneutic thinking, is also an important element of truth in the conception of the a priori as contingent and relative presuppositions of scientific research.

the particular experiences of life. If there are no a priori structures endowed with a determinate content, the mentioned pairs of concepts (phenomenon and noumenon, the point of view of the "I think" and that of the "thing-in-itself", the subject and the object of knowing) are all subsumed by the distinction and at the same time necessary complementarity of the functional irreducibility of the human person to any particular content, does not matter whether natural or cultural. In particular, if there are no categories endowed with their own content which the mind must necessarily make use of in knowing, there will not even be a phenomenon filtered, coloured or deformed by them, but there will be only the world of things, with their determinate characteristics, a world which is the object, from particular points of view, of an act of transcendental knowledge that must resolve itself in the multiplicity of the personal acts.

To express the same idea in other words, if the a priori is understood in a purely functional way, there can only be one 'thing' in our knowing that is a priori and that is irreducible to what is 'real', namely, the act by which the human person can in principle, as a functional subject, assume a critical detachment towards no matter what object. Understood in this way, therefore, the a priori cannot be manifold. Its unicity results from its empty, purely functional character, from its completely indeterminate character, which makes every multiplicity contradictory and incomprehensible. There must be only one a priori, which can be simply expressed as being a person. Here transcendental and personalistic approaches coincide perfectly: if the a priori is merely the act by which from time to time something is envisaged in a determinate way as an object, then its existence is resolved without residue in the individual persons who here and now think and actualize it.

From this point of view, one cannot accept Kalin's 1972 critique of transcendental arguments as *Gedankenexperimente*. As already mentioned, he rightly characterizes these arguments as (philosophical) TEs (Kalin 1972, p. 322). According to Kalin, given their strictly universal form, there is no way to show that transcendental principles (such as, e.g., the principle of cause according to which every possible object of experience has a cause), will remain valid in any future experience, unless one tacitly eliminates the possibility of encountering a contrary experience in the future, which however would make these principles analytic (cf. Kalin 1972, p. 327).

Now, this objection (echoing similar ideas of Lewis 1962 and Lovejoy 1967) holds only insofar as there still remains in Kant an a priori endowed with a content that should necessarily be found in any future experience.

On the other hand, the objection - which affects the empiricist conception of induction, understood as a logical procedure which, through the collection of particular cases, is supposed to exhaust the universality of a general law (cf. e.g. Cassirer 1910, 325-328) - has no value from a purely functional point of view of the a priori. From this point of view, the validity of transcendental principles can only be demonstrated negatively, showing that if one were to assume their negation, one would incur contradictions (or, in Kant's lexicon, antinomies).

Now, we are in a position to address the question of the nature and limits of philosophical TEs as transcendental arguments, because they have – if confronted with empirical-scientific TEs – the same relationship of unity and distinction that philosophy has with science. In the light of the above considerations, we should:

1) take into account Kant's considerations regarding the only difference in principle between philosophy and the particular sciences;
2) maintain a purely functional view of the a priori and consistently modify all the already mentioned Kantian theses that are linked to a material conception of the a priori.

As a result we may sketch four closely related main characteristics of philosophical TEs (and transcendental arguments).

First, we may extend two of the most important features of scientific to philosophical TEs: counterfactuality. TEs, whether scientific or philosophical, formulate questions concerning particular, paradigmatic cases to which they attempt to find an answer using counterfactual hypotheses. To formulate a thought experiment we must, at least implicitly, assume counterfactually that, given certain conditions, something would occur in a concrete situation.

Unlike scientific TEs, however, philosophical TEs have only an indirect relationship to the natural world, their first reference being the conceptual apparatus by means of which we describe the natural world or, more generally, the relationship (not only cognitive) to the world. To the extent that philosophical TEs reverse the direction of scientific inquiry, they are not questions posed to nature, but questions about our (historically changing) way of organizing, interpreting and evaluating what we call nature. We can thus summarize a first feature of TEs in philosophy:

1) philosophical TEs (or transcendental arguments) are exemplifications of counterfactual reasoning which express the same unlimited criticism that characterizes philosophy.

Second, because philosophy must find its objects outside itself, philosophical TEs – exactly as Kant's transcendental arguments or "experiments of pure reason" – bring to light some contradictions (or some ambiguities and confusions, from which contradictions can arise) only by referring, albeit indirectly, to what we call natural objects. For this reason they could not arise out of the vacuum of pure analysis: philosophical inconsistencies would not emerge without reference to concepts or laws already applied in experiences or in concrete and particular human evaluations. Hence a second characteristic of TEs in philosophy:

2) as far as their content is concerned, philosophical TEs (or transcendental arguments) depend entirely on information from "outside," that is, from common life experience and the empirical sciences, and it is this content that, albeit with respect to the conditions of possibility, grounds the criterion of their internal consistency.

Third, apart from reversing the direction of inquiry, there is no difference between the particular methods adopted in scientific reasoning and philosophical TEs (or transcendental arguments). Regarding these methods, the same considerations made in the case of the relationship between philosophy and experimental science apply here. The reversal of the direction of scientific inquiry does not mean that there is a difference in principle between the particular methods adopted in scientific and philosophical TEs. In discussing the conclusiveness of a philosophical TE, we can neither rely on a particular domain of objects, comparable to the objects of the various particular sciences, nor on a multiplicity of methods distinct from those of science. Given an a priori empty of content, in either case, in order to understand and then verify the truth value (or reliability, or whatever other cognitive or practical value one prefers) of a TE, we must take a genetic-reconstructive attitude and retrace, reconstruct, appropriate and evaluate for ourselves the steps that led to the assertion of that conclusion, or, in other words, the reasons why it had and still has to be accepted.

Hence a third feature of TEs in philosophy:

3) Apart from reversing the direction of inquiry, there is no difference in principle between the particular reasoning methods adopted in scientific and philosophical TEs or transcendental arguments (any such difference could only be based on a material conception of the a priori). To say the same thing from another point of view: reason-giving is the sole universal method of human reasoning.

Fourth, if there are no a priori or transcendental structures of the reason, neither philosophy nor philosophical TEs can have the task of describing or bringing to light such structures. Moreover, if the a priori is understood in a purely functional way, there can only be one 'thing' in our knowing that is a priori and that is irreducible to what is 'real', namely, the act by which the human reason can in principle assume a critical detachment towards any object. In other words, the a priori cannot be manifold. There must be only one a priori, which can be expressed as the pure function of being a person, i.e., as the function of being someway irreducible to any particular object. Here transcendental and personalistic approach coincide perfectly: if the a priori is merely the act by which from time to time something is envisaged as an object, then its existence is resolved without residue in the individual persons who, thinking in the here and now, actualize it.

We have argued that philosophical TEs, qua transcendental arguments, serve the purpose of demonstrating contradictions or impossibilities. From the personalist point of view, all the philosophical (transcendental) contradictions of the human reason with itself are different ways of obscuring the distinction in principle between the human person and any possible object (or, to use a famous Heideggerian expression, of thinking the person according to the model of the thing). In this sense, *the task of any philosophical TE (or transcendental argument) is always only critical or negative*, since it consists in demonstrating that the human person cannot be adequately described either through a priori rational structures or through a posteriori empirical structures. The negativity of this task does not, of course, detract from its complexity or inexhaustibility, since there are as many ways of entifying the irreducibility of the person to a particular (empirical or rational) reality as there are (empirical or rational) concepts through which we can describe it.

Hence a fourth (and perhaps the most general) feature of TEs in philosophy:

4) if there are no a priori or transcendental structures of the self or person, neither philosophy nor philosophical TEs can have the task of describing or bringing to light such structures, so that their task will always be only critical, negative, or, in a sense, Socratic: it consists in denouncing the contradictoriness of all attempts to reduce the human person to a multiplicity of a priori structures (which in the last analysis are always the hypostatisation of properties derived from empirical reality, whether natural or cultural).

Conclusion

In this paper, starting from a historical reconstruction of the Kantian concept of "experiments of reason," I have attempted to outline a theory of TEs in philosophy as transcendental arguments that is, at the same time, neo-Kantian and naturalistic. The idea of Kant's a priori, when understood in a coherently functional way, makes it possible to reconcile the thesis of a principled difference between scientific and philosophical TEs (or scientific and transcendental arguments) with the assumption of a methodological naturalism that admits of no difference between the particular methods of science and philosophy. The resulting conception of TEs in philosophy (or transcendental arguments) can be summarized as follows:

1) Philosophical TEs (or transcendental arguments) are counterfactual reasoning that reflects the same unlimited criticism of philosophy in general;
2) With regard to their content, they depend entirely on information from 'outside', i.e. from common sense and the empirical sciences, and the criterion of their internal coherence cannot sever all (though indirect) connections with this content;
3) Apart from reversing the direction of inquiry, there is no difference in principle between the methods of reasoning adopted in scientific and philosophical TEs (any such difference could only be based on a material conception of the a priori). Reason-giving is the sole universal method of human reasoning.
4) If there are no a priori or transcendental structures of the self or person, neither philosophy nor its TEs can have the task of describing or bringing to light such structures. Their task is only critical or negative. It consists in demonstrating again and again the contradictions in which our reason is involved when trying to exhaust the transcendental character of the human person with a priori structures.

References

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

- Buzzoni M. 2011. Kant und das Gedankenexperiment. Deutsche Zeitschrift für Philosophie 59, pp. 93-107.
- Buzzoni M. 2013. On Thought Experiments and the Kantian A Priori in the Natural Sciences: A Reply to Yiftach. Epistemologia 23, pp. 277-293.
- Buzzoni M. 2018. 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.
- Buzzoni M. 2019. Thought Experiments in Philosophy: A Neo-Kantian and Experimentalist Point of View. Topoi 38, pp. 771-779.
- Buzzoni M. 2021. The JanusFaced Nature of Philosophy of Science: Eleven Theses. Axiomathes 31, pp. 743–762.
- Cassirer E. 1910: Substanzbegriff und Funktionsbegriff. Untersuchungen über die Grundfragen der
- Fehige Y. 2012. 'Experiments of pure reason': Kantianism and thought experiments in science. Epistemologia 35, pp. 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, pp. 55-73.
- Fehige Y. and Stuart M.T. 2014. On the origins of the philosophy of thought experiments: The forerun. Perspectives on Science 22, pp. 179-220.
- Friedman M. 1992. Kant and the Exact Sciences, Harvard University Press, Cambridge.
- 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, pp. 62-83.
- Helms E. 2022. Ørsted, Mach, and the history of 'thought experiment'. British Journal for the History of Philosophy. 30(5), pp. 837–858.
- Illies C.F.R. 2007. Orientierung durch Universalisierung: Der Kategorische Imperativ als Test für die Moralität von Maximen. Kant-Studien 98, pp. 306–328.

- Jacobi F.H. 1815[1994]. David Hume über den Glauben, oder Idealismus und Realismus. Ein Gespräch. Beilage. Über den transzendentalen Idealismus. In: Werke, Bd. II, Fleischer, Leipzig, pp. 289-310. Quotations from the English translation by George di Giovanni, The Main Philosophical Writings and the Novel Allwill, McGill-Queen's University Press, Montreal 1994.
- Kalin M.G. 1972. Kant's Transcendental Arguments as Gedankenexperimente. Kant-Studien 63, pp. 289-328.
- Kant I. 1785[1997]. Grundlegung zur Metaphysik der Sitten. Kant I. 1900–, vol. 4. Engl. Transl. by Lewis White Beck, Foundations of the Metaphysics of Morals and What is Enlightenment?, Prentice-Hall, Upper Saddle River (New Jersey), 2nd edition, 1997.
- Kant I. 1786 [2004]. Metaphysische Anfangsgründe der Naturwissenschaft (MAN). Kant I. 1900–, vol. 4. Transl. and ed. by M. Friedman, Metaphysical Foundations of Natural Science. Cambridge University Press, Cambridge 2004.
- Kant I. 1787[1929]. Kritik der reinen Vernunft. In Kant I. 1900–, vol. 3. Engl. Transl. by Norman Kemp Smith, Critique of Pure Reason, Macmillan & Co., Limited, London 1929.
- Kant I. 1787[1998]. Kritik der reinen Vernunft. In Kant I. 1900–, vol. 3. Engl. Transl. by P. Guyer & J.W. Wood, Critique of Pure Reason, Cambridge University Press, Cambridge 1998.
- Kant I 1788[2015]. Kritik der praktischen Vernunft. Kant I. 1900–, vol. 5. Engl. Transl. by M. Gregor (rev. by A. Reath), Critique of Practical Reason, Cambridge University Press, Cambridge, 2015.
- Kant I. 1795[2006]. Zum ewigen Frieden: Ein philosophischer Entwurf. Kant I. 1900–, vol. 8, Engl. transl. by David L. Colclasure, Toward Perpetual Peace: A Philosophical Sketch, Cambridge University Press, Cambridge 2006.
- Kant I. 1800[1992]. Logik (ed. by Gottlob Benjamin Jäsche). Kant I. 1900–, vol. 9, English Transl. by J.M. Young, The Jäsche Logik. In: Lectures on Logic, part IV. Cambridge University Press, Cambridge 1992, pp. 527-629.
- Kant I. 1900–. Kant's Gesammelte Schriften, Königlich Preußische Akademie der Wissenschaften (Akademie-Ausgabe, AA). Reimer/de Gruyter, Berlin/New York.

- Kühne U. 2005. Die Methode des Gedankenexperiments. Suhrkamp, Frankfurt a.M.
- Lewis C.I. 1962. An Analysis of Knowledge and Valuation. Open Court, La Salle (Ill.).
- Lichtenberg G. Ch. 1968[1994]. *Schriften und Briefe. Erster Band: Sudelbücher I* (ed. by W. Promies), München/Wien, Carl Hanser. Quotations are from the 1994 edition.
- Lichtenberg G. Ch. 1971[1994]. *Schriften und Briefe. Zweiter Band: Sudelbücher II, Materialhefte, Tagebücher* (ed. by W. Promies), München/Wien, Carl Hanser. Quotations are from the fifth 1994 edition.
- Lipton P. 2003. Kant on wheels. Social Epistemology: A Journal oft Knowledge, Culture and Policy, 17, pp. 215-219.
- Lovejoy A.O. 1967. Kant's Antithesis of Dogmatism and Criticism, in: M. S. Gram (ed.), Kant: Disputed Questions, Chicago, 105-130.
- Mach E. 1903[1914]. Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen. Fischer, Jena. Quotations are from the 6th ed., 1914. Engl. Transl. by C.M. Williams, rev. by S. Waterloo on the 5th German edition, The Analysis of Sensations. Open Court, London 1914.
- Mach E. 1905[1976]. *Erkenntnis und Irrtum*, Barth, Leipzig, 1905, Engl. Transl. from the 5th edition, 1926 by Thomas J. McCormack (Chapters xxi and xxii) and Paul Foulkes (all other material), Reidel, Dordrecht/Boston 1976, pp. 134-147.
- Ørsted H.C. 1811 [1920]. Forste Indledning til den almindelige Naturlwre, in Naturvidenskabelige Skrifter, Vol. I-III, ed. by K. Meier, Host and Son, Copenhagen 1920, pp. 151-190. English translation: First introduction to general physics: A prospectus of lectures in this science, in Selected Scientific Works of Hans Christian Ørsted, pp. 282-309, translated and edited by K. Jelved, A. D. Jackson, and 0. Knudsen, Princeton, University Press Princeton.
- Ørsted H.C. 1822. Über Geist und Studium der allgemeinen Naturlehre. Gehlens Journal für Chemie und Physik 36, Nr. 4, pp. 458-488.

- Ørsted H.C. 1802. Ideen zu einer neuen Architektonik der Naturmetaphysik, nebst Bemerkungen über einzelne Theile derselben. Berlin.
- Popper K.R. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*, Routledge and Kegan Paul, London; 4th ed. 1972 (quotations are from this edition).
- Reichenbach H. 1920[1965]. Relativitätstheorie und Erkenntnis a priori, Springer, Berlin. Engl. Transl. by M. Reichenbach, The Theory of Relativity and A Priori Knowledge. University of California Press, Berkeley 1965.
- Schlick M. 1932 [1969]. Gibt es eM materiales Apriori? Wissenschaftlicher Jahresbericht der Philosophischen Gesellschaft an der Universitelt zu Wien — Ortsgruppe Wien der Kant-Gesellschaft fur das Vereinsjahr 1931/1932. Verlag der Philosophischen Gesellschaft, Wien. Quotations are from M. Schlick, Gesammelte Aufsatze 1926-1936. Olms, Hildesheim: 1969, pp. 19-30.
- Schmid J. 2020. The Experiment of Reason: Kant, Maimon, and Schelling on Metaphilosophy-First. Doctoral dissertation, ETH Zurich.
- Shi-Hyong K. 2008. Bacon und Kant: Ein erkenntnistheoretischer Vergleich Zwischen dem 'Novum Organum' und der 'Kritik der reinen Vernunft'. de Gruyter, Frankfurt.
- Stefanini L 1950. Metafisica della persona. Liviana, Padova.
- Tanona S. 2000. The Anticipation of Necessity: Kant on Kepler's Laws and Universal Gravitation. Philosophy of Science 67, pp. 421-443.
- Westphal K.R. 2003. Epistemic reflections and cognitive reference in Kant's transcendental response to skepticism. Kant-Studien 94, pp. 135-171.
- Westphal K.R. 2004. Kant's Transcendental Proof of Realism. Cambridge University Press, Cambridge.