Croatian Journal of Philosophy
Vol. XXIV, No. 71, 2024
https://doi.org/10.52685/cjp.24.71.4
Received: September 19, 2022
Accepted: June 26, 2024

The Mystery of Intuition in Einstein's Thought Experiments

MARKO GRBA
University of Rijeka, Rijeka, Croatia

The role of intuition in understanding in general and in scientific understanding in particular is still very much a subject of a lively philosophical discussion. The role of intuition in thought experimenting is much disputed in its own right, and the arguments range from those that deny any substantial role of intuition in the final inference that the thought experiment is meant to illustrate (eg. Norton or Williamson) to the pivotal role some form of intuition might play (eg. Brown or Miščević). So far, mostly Platonists were defenders of intuition, but in his recent book, Miščević takes on a formidable task to mount a defense of intuition as seen from a naturalist-evolutionist point of view and within his mental-modelling approach to thought experiments. I will, while acclaiming certain – and considerable – merits of his approach, nevertheless, insist that certain aspects of intuitive comprehending as it is meant to be going on in the process of thought experimenting remains inexplicable in the naturalist scheme such as Miščević's. The more mysterious (not to say Platonist) aspects of intuition will, hopefully, be revealed through the analyses of the two very famous thought experiments of Einstein which also figure quite importantly in his scientific opus. I will also have something to say about a few related problems as addressed by Miščević in his book regarding the description of thought experiment and more general imaginative enactments in thought, as well as on whether there is an essential difference between scientific (primarily physical) and metaphysical thought experiments and other thought experiments or related modes of thinking.

Keywords: Einstein; thought experiments; intuition; Miščević.

1. *The merits of Miščević's approach*

Miščević's new book on *Thought Experiments* (2022) is a most welcome addition to the growing literature on an important aspect of thinking in the natural sciences but also, more broadly, in theoretical and practical philosophy. The book is unique in being intended as a broad as possible an account of different theories of thought experiments (TEs) on offer in philosophical literature as well as of other related modes of thinking, from metaphysical TEs (Descartes' demon) to literary fiction (SF-stories for example), political utopias or dystopias, or even religious meditations (for example Ignacio Loyola's). For the whole lot of these mental modelling schemes which he, following Ernst Mach, sees as congenial, Miščević proposes a most ingenious phrase of *imaginative enactment in thought* (IET) (2022: 11). Thought experiment is then seen more specifically in the following manner:

> A typical TE starts with a design, which involves the determination of the goal(s) in the thesis/theory to be tested, and the construction of a *scenario to be considered*. It then proceeds with the presentation of the scenario thus constructed to the experimental subject, either the author of the scenario, or an interlocutor. In the later situation, the testing is done independently of the author: she is supposed to sit and wait for the verdict from the interlocutors, i.e. experimental subject's '*laboratory of the mind*'. On the side of the subject, the experiment starts with understanding of the proposed scenario, and then continues with the typically imaginative contemplation of it. Some reasoning might intervene. If all goes well, the subject ends with a verdict concerning the thesis/theory to be tested. Usually, in her mind it is presented to her as an invitation to believe or directly as a belief, most often seeming obvious and compelling. Such states (invitations to believe, or immediate beliefs) have been traditionally described as '*intuitions*'; they are often likened to similar states concerning mathematical insights or obviously looking linguistic judgments on sentences in subject's native language. Once the verdict is achieved, it can be and often is compared with results of other scenarios in the vicinity, or other versions of roughly the same scenario. Finally, interesting and provocative verdicts are normally being brought to comparison with items of knowledge or widely accepted beliefs. If they clash, the arduous task of balancing is required, in which the particular verdict might win (as has historically been the case with Galileo's verdict on falling bodies), or, alternatively, the established knowledge might, or, thirdly, some compromise is made. *The result is usually described as* '*reflective equilibrium*' (2022: 14, my italics)

Miščević is tying in one finely knit fabric a vast body of views and analyses found in literature, such as James R. Brown's (1991/2005) idea of a *laboratory of the mind* as the scene of thought experimenting, or John Rawls' *reflective equilibrium* of judgments, and presenting to the reader a unified picture of the whole realm of modes of thinking which have been used by various authors and to various purposes for millennia under one name and one guiding principle. As I take it, this guiding principle is to see how scientists, philosophers and authors are generally arriving at their ideas, more or less revolutionary, relying on their

intuitions and employing their imagination, perhaps (at least at times) more than their logical reasoning. This fits very well with what Einstein said about the respective role of intuition/imagination and logic in the context of discovery of the fundamental laws of nature:

> The supreme task of the physicist is to arrive at those universal elementary laws from which the cosmos can be built up by pure deduction. There is no logical path to these laws; only intuition, resting on sympathetic understanding of experience, can reach them. (Einstein 1919: 226)

Connected to the idea of a unifying approach to studying different modes of thinking in as diverse fields as physics and political theory, we might speculate on why certain ideas were historically seen as (more) revolutionary than others, say, Copernican revolution in astronomy as more revolutionary than Plato's epistemology, or on a par with the ideas of the French revolution. Could it not be that many a time ideas and, indeed, the values of ideas were judged more on the merits of their practical application, or potential for such an application, rather than on their intrinsic (theoretical) value? Although, Miščević is not likely (based on what I know form our conversations) to agree with Einstein's deductivist position (as espoused in Brown 1991/2005: 112–121) as to the methodology of science, or embrace a Platonist epistemology, nevertheless, his account is potentially broad enough to accommodate even such widely differing positions on the epistemological spectrum.

Reading Miščević's book, one could gain an impression that his intention was to write a sort of a guidebook on how to conduct thought experiments, given the detailed analysis of their structure or the breadth of examples and references. In many respects, I would say, one would not be amiss to take advice from this book. However, one must always take it with a pinch of salt, especially when it comes to how to understand intuition as such and what exactly it takes to reach a conclusion from a thought experiment. These are the issues I will now take on in the next two sections basing the discussion on two most consequential thought experiments of Einstein.

2. *Two conundrums regarding physical TEs (applicable to other scientific TEs)*

Miščević's account of IETs (2022: 67–68), includes model building, thought-experimenting and intuition-producing. Regarding the TEs as a subspecies of IET he demands that they are scenario-based rather than inference-based, that they produce intuitions as their final products in a process of mental modelling where various highly particularised scenarios are played as in front of our eyes and in the creation of which imagination of the experimenter (speaker/interlocutor) has a central part to play. He insists that such scenarios have both cognitive and justificatory role in TEs, whereas inference plays a subordinate role. Furthermore, he gives a pivotal role to intuition as having to do

with the external referential domain, not merely concepts. For him this intuitioning is largely innate and related to a specific competence(s) of the brain along the lines of the standard Chomsky's proposal. Miščević, however, goes a step further and generalises the specific lingusitic competence to other crucial competences when it comes to understanding and dealing with the world (such as spatial, temporal, numerical etc.). He does not claim that what intuitions share is primarily the underlying structure(s) in as much as it is the manner of representation. These competencies are ultimately regarded in an evolutionary-adaptationist way. This approach to understadning TEs, and more generally IETs, he names the *Moderate Voice of Competence* proposal (MoVoC):

> So, we now have the minimal necessary elements to formulate a proposal concerning the nature of intuitions and TEs producing them. I have called it Moderate voice of competence view ('MoVoC' for short). It starts from the admission that there are intuitions-dispositions and judgments, which form a distinct group of phenomena, and there is the intuition-capacity, the capacity to use our imaginative and judgmental competencies in an off-line fashion. It is the voice of competence, most often discreet. Intuitional data are thus the minimal 'products' of tentative production – linguistic, philosophical, moral or mathematical – by naïve thinker (or speaker-listener) and not their opinions about the data. The data involve no theory and very little proto-theory. Although there might be admixtures of guesswork in the conscious production of data, these are routinely weaned out by linguists. As against predominantly conceptualist understanding of TEs and intuitions (Peacocke, Boghossian) it claims that intuitions are concerned with their external objects, the domain of items and facts, rather than with concepts. Concepts often play a role in the process, but they are not the object of intuitions, and their role is subordinate to the role played by the external referential domain. (2022: 67)

Although I agree with the general framework, especially with putting the stress on the key part the imagination plays in TEs, the view that imagined scenarios have both cognitive and justificatory role as well as with assigning the intuition an external referential domain, as all these features seem to me prominent in scientific TEs, especially Galilei's and Einstein's, I would be somewhat hesitant in committing to the very narrow evolutionist-adaptationist account of the origin of intuition-capacity or the thought process as such. Given, first, that what we know on these matters is still mostly informed by research from psychology rather than neuro-science which is both more "naturalistic" as well as more accurate in its measurements, and hence conclusions than psychology will ever be, and yet does not really give us much to muse about at the present state of development. (One may wonder whether it ever will, given that some of the problems are related to the problem of *qualia*, which is notoriuosly difficult to solve from any point of view). Furthermore, even if we assumed that valuable intuitions which will have some bearing for the understanding of the world might be arising in special mental capacities pertaining to specific brain region(s), we could ask why the more sophisticated intuitions do not arise much

more often, as I believe it could be agreed on that the insights of the kind Galilei or, even better, Einstein had, arise in others even in a span of a century. If this were the case, our science would have been on a much more developed stage by now?

In sum, the first conundrum I see unresolved in Miščević's work so far (as in most of other authors) and not much commented on either, would be the origin or, (I guess) in Miščević's case, the mechanism of generating the more sophisticated intuitions. If I understood him well, in case of scientific (physical) TEs, the proposed source of these ideas to be identified with some sort of folk science/physics (as modelled on folk psychology, which in my above described view already makes it a problematic idea), the concepts/intuitions which are then clashed with the *accumulated wisdom* of the ages (2022: 64) and tempered by new (real) experimental data, simply will not do. Even Miščević agrees that the ideas of the folk sciences are usually fallible as: "Our innate geometry might be false, our possibly innate folk-physics certainly is" (2022: 65). Not to mention that it is hard to sometimes even formulate what the folk-scientific ideas would even be, say in the case of chemistry (as most humans do not perform that many relevant real chemical experiments to acquire a significant body of observations which could then be conceptualised in any meaningful way). In the case of physics, the situation should by no means be underestimated, given that almost all fundamental physics concepts are to a high degree sophisticated. There is nothing obvious or simple in any of the concepts we use in, say, Newtonian mechanics: such concepts as speed or acceleration already have both a scalar and a vector representation (which obviously assumes the knowledge of a sort of vector algebra); the ideas of motion, continuity of space and time or matter are debated since the pre-Socratic philosophers and still mostly unresolved. Galilei and Newton came to their fundamental postulates of the science of mechanics by a combination of highly sophisticated abstraction and pure guesswork (with a little bit of experimentation where the limitations of air-resistance or friction allowed it). But the real challenge is to try to account for any of the sublime thought experiments of Einstein by way of relating his new ideas to some form of folk physics, or folk mathematics, especially for the more consequential of his TEs. Some of the challenges will be presented shortly.

The second conundrum I see unaddressed is how exactly does the inference come about from the thought experiment as this again is by no means obvious. Especially so in the case of the sophisticated TEs like Einstein's. Miščević, in my view, provides persuasive arguments in favour of an intuitionistic view of TEs (and IETs) as opposed to inferentialist or conceptualist views, but I would like to have seen this relation of inference to the scenario of the thought experiment described in more detail as this is where the real trouble begins when it comes to interpreting the TEs or ascribing any value to them in the context of,

say, theory building such as was Einstein's regular practice. Norton's famous account (1991) skillfully avoids asking part of the epistemological question about the origin of this relation. He is only interested in spelling out the logical part as he much later admitted in a response to a criticism (Norton 2021: 125–126). The origin of the relation for Norton is resolved by assumption of identity, where *thought experiments are simply picturesque arguments*. But what about Einstein's words as quoted above insisting that: "There is no logical path to these laws; only intuition, resting on sympathetic understanding of experience, can reach them?"

3. *Process of discovery and process of justification in the case of Einstein*

I would argue that the two most consequential questions to be answered when it comes to interpreting the results of a TE are: *Which idea(s) do(es) the explaining?* and *How does one arrive at the idea(s)?* The second question is not only relevant in the context of discovery but could also be in the context of justification, to use the famous distinction by Reichenbach. It could happen, namely, that the path to discovery (the heuristics if one prefers) might be of a significance also as steps of justification, which is how Einstein often argued when trying to give an account of justification of his theories (including both special and general theory of relativity), as Norton convincingly argued in (1995). Most often (definitely in the case of Einstein's TEs) the idea(s) that actually serve(s) as explanans is/are quite subtle and unexpected (so it would appear that there is not much there in terms of folk physics, as Miščević demands it), to the point of being of inexplicable origin, or at least origin hard to trace. Two very famous TEs will be used to illustrate: Einstein's elevator and his light momentum TE with the help of which he derived $E = mc^2$. But before those, a word or two on the comparison of Norton's views to Miščević's as I believe some interesting thoughts might emerge.

Even a rationalist and inferentialist, when it comes to analysing the origin or structure of a TE, like Norton, admits (1995: 63) that a rationalistic account of the discoveries (and thought experiments) of Einstein leaves room for *arational*, in Norton's own words, elements and, as he puts it, *perhaps even Einstein's "free inventions of a human mind*." But the key question here, surely, is how much exactly in genius's (like Eintein's) process of discovery is rationally accountable and how much remains perhaps forever inexplicable, at least by a rationalist analysis? Of course, this is very hard to establish. However, it does matter a great deal for the following reasons.

First, if key ideas came to Einstein in some sort of an epiphany (much like to the mathematician Ramanujan in a dream, according to his own recollection passed to G. H. Hardy. This, of course, annoyed rationalistic and logical mind like Hardy's, especially given Ramanu-

jan's insistence that he needed no proofs for the mathematical propositions thus revealed). They were presumably unique, or at least quite specific to one mind, that of Einstein's. This means that it would be a gross oversimplification to claim that the subsequent rational analysis of the origins of these ideas is possible or even useful. To the contrary, one could, after reading such analysis, acquire a completely distorted picture of the real process (if there was any) and assume that if only one would follow the steps of the rational analysis, one could repeat the same kind of discovery, or achieve the discovery of the same calibre as some of Einstein's discoveries. Now, I am not arguing that no rational analysis is ever possible – far from it – but simply that more space and a more of an open mind should be left to the possibility of the contrary. The contrary could then be seen as either a Platonic insight of a sort, or a naturalistically founded intuition as understood by Miščević and described above. This would then be an argument in favour of Miščević's conception of the process of discovery in TE, but also a defense of Miščević's account of nature and value of a TE as against Norton's. However, I have an issue with Miščević's account of the discovery process as too narrow in not allowing for anything but a naturalistically understood intuition. But how then to account for the rarity of such deep insights as Einstein's? Surely, if evolution has programmed us for such deep thinking, then it must have programmed more of us, proportionally many more then the history of science would allow for (by which I mean the history of those ideas in science that have proved fruitful especially when it comes to the predictive power of natural sciences)! On the contrary it would appear, that Einstein was quite unique in his way of thinking as well as discovering.

The second objection to a thoroughgoing rationalist analysis of the type of Norton's is to my mind even more serious, if not even deeper. Namely, the objection that follows from the point raised by Einstein as quoted above, that *only intuition, resting on a sympathetic understanding of experience,* can reach deep insight into the fundamental laws of nature, which, as I claimed at the beginning of this paper would go in favour of Miščević, but not necessarily of his naturalism. For Einstein surely knew what he was talking about and his *dictum* was inspired by his own experience of working for decades at the forefront of research in foundations of physics, from particle physics to cosmology, and so his emphasis on intuiton as *opposed* to logic must have had some grounding in observing his own process of discovery. This insistence would appear to agree well with Leibniz's view of reasons of the world of physical phenomena which never necessitate but only incline (Russell 1937/1992: ch. 3), meaning that the connection of no two ideas in physics is logically necessary, hence it is not possible to discern such a connection by applying pure logic. It would be interesting to know what Miščević's thoughts were on this aspect of the problem of acquiring knowledge about the physical world.

Finally, I arrive at my perhaps most controversial point, which I am inevitably led to, especially after having spent some time assessing the merits and demerits of various accounts of how Einstein came to discover his general theory of relativity and this associated elevator thought experiment. My point can again be well posed as against Norton's claim in the above referred paper of 1995 (62–63) to the effect that the better the rationalistic reconstruction of the process of discovery is, the less mystifying the process appears and, consequentially, the more likely the steps of the process of discovery are to be also seen as the steps in justification or explanatory process, if this can at all be achieved (as Norton, I believe, justifiably claims Einstein himself was in the habit of doing, at least when it came to the theories of principle, as he called them[1]). The point is that if Norton is right that sometimes (at least in the case of some of the steps along the path of discovery of Einstein's theories of relativity) the process (or parts of the process) of discovery can be supplanted for the process of justification, so heuristics could be supplanted for logic. If so we should be extremely observant as to the details involved as is best seen in the case of Einstein using the so-called *equivalence principle* in discovering the general relativity which will now be briefly described as some of the best available accounts in the literature. The claim I will be led to is that in the case of using the equivalence principle (or principles as Einstein actually changed the meaning of his postulate on several occasions), and as most famously exemplified in his elevator TE, Einstein was in the end making up a just-so-story rather than presenting a genuinely valid argument or operational TE to support his quest for general relativity, although not doing it consciously, at least not at all times.

4. *Einstein's elevator TE and the equivalence principle as idée fixe*

Soon after completion of his special theory of relativity, which was Einstein's response to the most pertinent issue of the day, namely, the conflict between Newtonian mechanics which embodied the principle of relativity of all motions and Maxwell-Lorentz electrodynamics which seemed to suggest the independence of the speed of light of any source or direction of motion, Einstein embarked on an even more ambitious

[1] The theories of principle, as opposed to constructive theories, according to Einstein, are those that are founded on a minimal number of preferably empirically suggested basic principles (axioms) which do not assume anything about the structure of the material world, only state some universal properties of natural processes or their theoretical representations which then have to be cast into mathematical form (Einstein 1919: 228; Brown 1991/2005: 103–105). An example of such a theory of principle, after which Einstein modelled his theories of relativity, is classical thermodynamics with its main principles being the laws of thermodynamics (viz. the impossibility of building the perepetuum mobile of either the first (1st law of thermodynamics!) or the second kind (2nd law!).

quest – to generalise his theory. Although the special theory of relativity was a remarkable achievement in its own right, especially given the minimalist nature of its structure and the scarcity of experimental evidence at the time (early 20th century), it was a theory of a limited domain of application, applying only to systems in uniform non-accelerated motion, to motion of the so-called *inertial reference frames*. But Einstein sensed, rightly as it turned out, that the basic structure of the theory, what in mathematical terms would amount to invariants of motion with respect to a certain group of symmetry transformations and in physical terms would have implications for the way we represent spatial and temporal relations between phenomena, held a much bigger promise. However, this was initially only a pretty vague impression, although strongly present in his mind. For Einstein in his twenties (when he was developing his special theory) was not yet a fully trained mathematical physicist as he was to become during the work on his generalised theory, for which he had to develop a mastery of the latest developments in then-contemporary mathematics (such as the absolute differential calculus, or tensor calculus, of Ricci, Levi-Civita and Cartan). Indeed, he had to learn to appreciate the fact that further advances in ever more abstract theories of fundamental physics come (perhaps) exclusively at a very high price in terms of the mathematical knowledge requisite in their development (see eg. Norton 1995: 61–62).

As late as 1914, Einstein still had doubts about whether he should follow the path of mathematical simplicity and elegance or carry on in his familiar way through direct physical insight. Different authors see these internal quibbles as a consequence of Einstein before 1915 still being naive to abstract mathematics of his day, but one could, at least with a hindsight, see in these an originality of approach to physics as Einstein's characteristics, as perhaps one of only very few physicists or scientists of his statue. Namely, Einstein was hesitant to adopt the predominantly mathematical approach to problem-solving in the realm of physics as he was genuinely baffled by the ever-increasing demands in terms of the level of mathematical sophistication, which is usually accompanied by an appropriate increase in the level of abstraction, with every new and more subtle problem in physics In effect, Einstein was overawed by the ramifications of the relation between mathematics and physics. Rightly so! As I would dare say, whoever takes the complexity of this relation lightly usually pays the price of losing the compass as to what could exist in reality but which is not revealed in mathematics alone. And there would appear to be such an element in at least every mature physics theory. So Einstein was not wrong in being *prima facie* suspicious towards giving mathematics the predominant role in guiding the research in physics/science, but only extremely cautious. At his expense, as it ultimately turned out and is well known, since by 1916 he was able to complete the general theory of relativity which was actually a new theory of gravity, only by fully adopting all the sophisticated

mathematics which he could learn from his mathematician friends, some of which were among the greatest mathematical geniuses of all times (like Tullio Levi-Civita, Felix Klein, David Hilbert, Emmy Noether, Hermann Weyl and Elie Cartan, more or less in historical order of appearance in Einstein's professional life). The question, however, remains, what was Einstein relying on, if not highly abstract mathematical techniques, in deriving his conclusions in physics? The answer is also well known: primarily thought experiments!

As early as 1907, Einstein thought about generalizing his theory of relativity since he was naturally dissatisfied with it being applicable only to a very narrow domain of uniform inertial motion and was looking to extend the domain of application of, first and foremost, the relativity principle to all the relative motions. Einstein's original train of thought might have looked like this (see eg. Janssen 2014 with my own insertions here and there): given that in special theory no one's frame of reference could be thought as absolute (as one cannot prove its existence by, say, observing motion relatively to this frame) and that all the inertial motions are hence only considered as relative, one should expect that all the reference frames are equivalent (including the accelerating ones) and so any one could be deemed as at rest for a given observer. In essence it is to find the most general form of the laws of nature, independent of the choice of the coordinate frame. The task seems meaningful enough, indeed, something to be desired, as if there is no favoured frame of reference, then surely the fundamental laws being universally valid entails their mathematical formulation being coordinate-independent. The trouble is that the effects of non-inertial motion (such as tidal forces due to gravitating masses) are discernible for all the observers alike, whether moving with the frame or apart from it. At the time the only candidates for fundamental forces (to which all others would reduce in final account) were electrical, magnetic and gravitational forces. Since Maxwell showed electrical and magnetic forces to be two sides of the same unified electromagnetic force (or rather field), and given that magnetic force was shown by Einstein himself to be eliminable by a change of a reference frame, Einstein might have been inspired (we do not know this for sure!) to ponder upon the idea of somehow eliminating gravity as a force, or, rather, transforming gravity away and therefore transforming between inertial and non-inertial reference frames. Thereby, in the long run, perhaps achieving the transformation between any and all the reference frames as if any one of them could be at any moment seen as at rest. While still at the patent office in Bern, Einstein had, in his own words, *the happiest thought of his life* (as quoted in Janssen 2014: 174 and note 30): what if a man was falling with the elevator, would he not have the same experience as if in a state of weightlessness in a space without gravitational fields? By extension, an observer who is performing various observations in a stationary elevator in a gravitational field would have the same experi-

ence as the one who is moving in an elevator (in space free of gravitational fields) which is acted upon by a force different from gravity but acting so that inside the elevator everything appears as if there is no external force on the elevator but gravity. Does that not suggest a way to eliminate gravitational force and still have the gravitational effects?

As mentioned, Einstein's efforts towards generalization of his special theory started as early as 1907; at first by him trying to develop a special-relativistic account of Newtonian force of gravity, but that, he soon realised, was impossible given that Newtonian force assumed instantaneous action at a distance and special relativity implied relativity of simultaneity. Einstein struggled for several years with different versions of a special-relativistic theory of gravity as did some of his contemporaries (like Max Abraham or Gunnar Nordström) and managed to conclude that no such theory (either a (3+1)- or 4-dimensional) is possible. The interesting point from those early attempts is that Einstein used what he is to call the equivalence principle only later. At the time, this meant that Galilei's law of free fall holds, namely that all objects falling from the same height in a homogenous gravitational field fall at the same rate and that the vertical velocity of fall is independent of the horizontal component of motion, if there is such. The special-relativistic theories of gravity did not fulfill the second part of the statement (as well as the law of conservtion of energy which was by then taken as "sacrosanct" in physics). Now, Einstein was to relate to the Galilean law of fall another implication, namely, that of the equivalence of inertial and gravitational mass, which is a fact tacitly assumed in deriving the law of motion of a (point) mass under the influence of Newtonian force of gravity, which Newton too was aware of. But none of these on its own, or standing together, would enable Einstein to make any progress from special theory of relativity to the generalised form of any kind, as is clear from his elevator TE, which actually assumes much more, albeit not clearly expressed. Einstein, indeed, was aware of the limitations of merely coupling special relativity with the law of free fall, the equivalence of inertial and gravitational mass and having Newton's law of gravity as a limiting case of his new theory. And, yet, what else could he demand for his general theory to fulfil?

He played with various ideas, having a variable speed of light (sacrificing the second of the two postulates of special theory of relativity all to generalise the first), but all in vain, as it turned out. He then envisaged another of his thought experiments (Janssen 2014: 178-181), the rotating disk as a frame of reference. Wondering how an observer at the circumference would see the passage of a light beam sent from the observer at the centre of the disk towards the circumference, he concluded that although the light would travel the straight line path, it would not be perceived as such by the observer at the circumference due to difference in linear velocity of the two observers. Given that, after the elevator example, the rotating disk frame (with centripetal

force) is equivalent to a disk at rest with a centrifugal gravitational field acting on it, we are justified in concluding that gravity bends light, as indeed the final form of the general theory of relativity accurately predicts regardless of how bizarre either the conclusion in the first instance of the rotating disk or the transference of it to the disk at rest may at first seem. Now, the geometry of the rotating disk is meant to be Minkowskian as in special relativity, which enabled Einstein to make deductions, given its familiarity. The principle of equivalence (proclaiming the equivalence of effects as seen from the appropriate system in accelerated motion or the one at rest in a corresponding gravitational field) enables the transfer of deduction to another type of reference frame, namely the one in a gravitational field, hence giving Einstein an essential insight of the outline of the sought after general theory. It also may inspire, as indeed it might have inspired Einstein, according to Stachel (1989), to consider alternative geometries to Euclidean for the space-time structure of general theory (here Einstein would have considered contraction to measuring rods along the radius and circumference and from these deduced ratios of circumference to the radius of a rotating disk which might differ from 2π).

Even if Einstein has managed by, more or less, following the path described above to reach correct conclusions, there are certain conceptual problems which cannot be ignored, and, indeed, Einstein could not ignore them when discovering general relativity. The alleged equivalence which Einstein crucially relied upon in making his deductions valid for the case, first, of a homogeneous gravitational field, and then any gravitational field turns out to be extremely difficult to articulate, so much so that Einstein changed the meaning given to his principle on several occasions, after more than ten years finally reaching the mature form which aims to make any gravitational field as having only relative existence as one side of the so called *inertio-gravitational field*: "There is only an inertio-gravitational field that breaks down differently into inertial and gravitational components depending on the state of motion of the person making the call" (Janssen 2014: 178).

Now, the problem with this formulation, although Einstein claimed it was the key for discovering general relativity, was that it retroactively(!) sanctioned the inference of gravitational effects from accelerated frames with Minkowski space-time given that the very metric of Minkowski space-time would, according to mature equivalence principle, be taken as a particular inertio-gravitational field (Janssen 2014: 179). This means Einstein would have been able to make the deduction as found within the Minkowski frame valid in a frame at rest with gravity acting only *since* the metric of Minkowski space-time represents a form of gravity! There was never actually an equivalence between an accelerated frame and a frame at rest with gravity, but only between two types of frames with gravity. This could further be read as a curious case of a let's-pretend game (or just-so-story) that Einstein

was playing on himself for several years. Of course, it could only have worked if the metric can be identified with gravity, but here again we have several major issues to consider which Einstein was to become gradually aware of either through his own efforts or through constructive criticism from colleagues.

Einstein claimed for many years in his papers and correspondence with many authors/critics that gravitational fields should only have relative existence, that equivalence principle helped in a crucial way on his path towards the covariant field equations of general theory and that the general theory was a generalization of the invariant special theory of relativity in the sense of relativity of all reference frames and all motions. We now know that none of these actually holds, and that even the extent to which the equivalence principle really helped or hindered his research is not quite clear as seen from the writings of different authors (for instance Janssen in (2014) argues for more of a hindrance case, whereas Norton in (1985: 40) praises Einstein's use of equivalence principle *as one of the most beautiful of Einstein's insights*).

With respect to the claim of relative existence of gravitational fields (Janssen 2014: 178–179), it follows from Einstein's insistence that the gravitational field is to be related to the metric tensor not the Riemann curvature tensor of Einstein's field equations. This follows from his reinterpretation of special relativity theory as a theory of special case of gravitational field, namely the one generated by the Minkowski metric, which means that now even non-accelerated frames could be associated with a field. Thus any field could be thought of as related to a reference frame and transformable, therefore of relative existence. At first, and for a long time, Einstein thought this idea, coupled with what he in 1918 dubbed as Mach's principle,[2] could finally remove any trace of absolute motion from physics. In this he turned out to be wrong as the Dutch astronomer and mathematician Willem De Sitter managed to show after which Einstein gave up attempts at a Machian account of inertia, but not before introducing his (in)famous cosmological constant to the equations of gravitational field, which he later denounced as his *biggest blunder*. By 1954, in the final year of his life, Einstein wrote:

[2] As Einstein wrote to De Sitter in 1917 (quoted in Janssen 2014: 202): "It would be unsatisfactory, in my opinion, if a world without matter were possible. Rather, it should be the case that the $g_{\mu\nu}$-field is *fully determined by matter and cannot exist without the latter*. This is the core of what I mean by the requirement of the relativity of inertia." Which means that there is no field without matter which generates it and given that all motion is with respect to metric ($g_{\mu\nu}$), it is in actuality with respect to some constellation of masses as Mach originally conceived in response to Netwon's famous bucket experiment which had as aim proving the existence of absolute motion by example of rotation of a water inside a bucket even when the bucket does not move relative to the water, after being set in motion from rest with the initially still water and then stopped at the point when the water reaches the highest point of ascent of concave surface as against the walls of the container.

> In my view one should no longer speak of Mach's principle at all. It dates back to the time in which one thought that the 'ponderable bodies' are the only physically real entities and that all elements of the theory which are not completely determined by them should be avoided. (I am well aware of the fact that I myself was long influenced by this *idée fixe*). (Einstein to Felix Pirani, February 2nd 1954)

I claim that similar judgement could be passed about the equivalence principle and hence about what is usually claimed to be the gist of the elevator TE. Indeed, this judgement was passed by the leading relativist of the later generation, John L. Synge, and is a predominant view among the physicists working in the field of general relativity and cosmology (Janssen 2014: 178–179) given that the modern day criterion for the presence of a gravitational field is whether the curvature (not metric) tensor has non-vanishing components: "The Principle of Equivalence performed the essential office of midwife at the birth of general relativity. […] I suggest that the midwife be now buried with appropriate honours and the facts of absolute space-time faced" (Synge 1960: ix–x).

As Janssen explains in conclusion of his überblick of the genesis of general relativity (2014: 208), Einstein could indeed be said to have developed a theory which can be interpreted as seeing gravitational fields as relative, but definitely not all motions as relative. Also, the invariance of the covariant equations of general relativity which Einstein was for a while conflating with invariance of special relativity (as pointed out by Erich Kretschmann already in 1917 and by several authors ever since (Janssen 2014: 186–187)) thinking that there is a principle of relativity of motion related to the general theory as well. Furthermore, Einstein's hopes to make general theory into a Machian theory of gravitational field and inertia failed and Einstein, as we have seen, in the end gave up Machian notions. What, then, remains, is a new theory of gravity with absolute pseudo-Riemannian space-time, still with some vestiges of absolute motion and hard to trace genesis.

Next to the ideas of non-Euclidean geometry of space-time and covariance of the field equations which Einstein picked up through studying abstract mathematical theories and musing on whether these could have any consequence for the physical reality—much like Gauss once wondered whether the sum of the angles in a physical triangle (made of, say, light beams from different lanterns sufficiently far apart) is 180° or more, or less—there is one more idea which makes for a constant in his thinking throughout the process of discovering general relativity. This is the idea I mentioned at the beginning as I believe it was one of the earliest thoughts in this process, namely, the intuition that gravity should be eliminable as a force. Now, at the end of the discussion of the elevator TE and related equivalence principle, let us examine how credible this idea is. I think there are reasons to believe this is the idea the elevator TE was supposed to illustrate all along and it is the one constant that crops up again and again in Einstein's thinking after 1905.

At the time of the development of general relativity this idea, however bizzare, could have appeared reasonable enough to push forward, given in particular the analogy with how magnetic field can be eliminated in special realtivity. But what about electric force/field, can it be eliminated in the same way? Or, even better, what about nuclear forces, the strong and the weak, which could not only be said not to be eliminable but also could not even be conceived as classical fields? So one wonders whether Einstein would have hit on his covariant field equations, at least if he would have discovered them starting from the same originating ideas, if he already in the 1910s could have known of the other two fundamental forces? To conclude, the mixed messages we get from the elevator TE are just a sign of the more general cacophony that still remains when it comes to disentangling all the subtleties involved in discovering and justifying the general theory of relativity.

I would now go on to analyse messages of another of Einstein's famous TEs, this time with a more positive conclusion, and return at the end of the next section to the problems surrounding the Einsteinian justification process which could also arise as problems for a non-Platonist account such as Miščević's.

5. *Einstein's light momentum TE: deducing $E = m c^2$*

I believe the less discussed of Einstein's thought experiments, the light momentum TE, deserves perhaps the highest status. It would appear that Einstein regarded it highly too, as he developed versions of it virtually throughout his working life, from 1905 to 1946. The version presented here is Norton's adaptation of Einstein's 1946 and final rendering (Norton 2014: 96–98). Why would Einstein return on multiple occasions in the span of more than four decades to try to demonstrate that $E = m c^2$, or that energy and mass are equivalent? Each time trying to render his "proof" simpler and using fewer elements from parts of physics different from special theory of relativity. And what can be said about the nature of his proofs? Are they to be taken as sufficient *per se* to establish the relation, so as *a priori* (or mathematical) proofs, or should we still require experimental evidence that the relation holds (which we have by now obtained on many occasions from different type experiments)?

The answer to the first query would appear to be that Einstein wanted at least one quantitative result of his two main contributions to theoretical physics and science, in general, to be fully within grasp of even a high school student of physics, and I believe with the final version of his derivation (as presented by Norton in any case) he indeed succeeded, given the minimal requirements of knowledge of either physics theories or its experimental results (which Einstein anyway lists and none of which is too difficult to understand or at least appreciate in its significance for the derivation) as well as of the level of mathematical skill (basic high school vector algebra will suffice). It would

make perfect sense for Einstein to try to achieve such a derivation/argument, as in spite of his theories of relativity being quite abstract and conceptually extremely demanding (especially the general theory, as we have even if only partly seen in the previous section), not to mention the mathematical requirements they impose on the student, Einstein fostered a firm belief that the fundamental ideas of his physics, as indeed of all physics, can be expressed in simple terms (at least some of those ideas). But I believe there is yet another reason which he expressed perhaps most clearly in his famous 1933 Oxford lecture on the methods of theoretical physics:

> Our experience up to date justifies us in feeling sure that in Nature is actualised the ideal of mathematical simplicity. It is my conviction that pure mathematical construction enables us to discover the concepts and the laws connecting them which give us the key to the understanding of the phenomena of Nature. Experience can, of course, guide us in our choice of serviceable mathematical concepts; it cannot possibly be the source from which they are derived; experience, of course, remains the sole criterion of the serviceability of a mathematical construction for physics, but the truly creative principle resides in mathematics. In a certain sense, therefore, I hold it to be true that pure thought is competent to comprehend the real, as the ancients dreamed. (Einstein 1934: 163–169)

I mean, in particular, his emphasis that *pure thought is competent to comprehend the real, as the ancients dreamed,* and the identification of this thought with the (predominantly) mathematical process of discovery, which of course is primarily aprioristic, as is its justificatory process. By referring to the *ancients* (not the modern day philosophers!) Einstein is further underlying to which genealogy he as a thinker belongs, to the genealogy of Platonic thinkers (at least partly, given that Einstein did use different epistemologies in an opportunistic way depending on the needs of his science), those who *dream* that reality can be comprehended by pure thinking (where this is clearly not meant in a pejorative sense). To the latter testifies the opening phrase of the quoted paragraph: *in Nature is actualised the ideal of mathematical simplicity*. The experience Einstein here refers to is his own experience of work (by the time of his speech measured in a few decades) at the forefront of research in theoretical physics. To me it is also significant that he uses another of Leibnizian expressions about actualization of principles. Einstein as an avid reader in philosophy and, if not consciously a follower of Leibniz but explicitly a follower of Spinoza, with whom, as is well known, Leibniz had many points in common (not to mention that he went to study with him as a young men). So Einstein, in polishing his "proof" which he reached by pure thought (as no experimental evidence was available for it around 1905 and for years to come), like Spinoza was polishing his lenses, could be seen as trying to present a perfect proof of his general approach to physics and partly of his worldview.

Before we endeavour to answer the second query, let us examine the derivation. The process of light emission is seeen from two frames of reference, one at rest (S', with mass of the emitter m' and emitted energy in both directions $E'/2$) and one moving with velocity v perpendicularly to the direction of emission as seen from S'. Momentum of light is given from Maxwell's theory by $p = E/c$, and thus from the viewpoint of S' the momentum of light in each direction is $E'/(2c)$, whereas the new momentum from viewpoint of S is

$$(E/2c)(v/c)=1/2(E/c^2)v,$$

taking into account only the vertical portions which are a v/c fraction of the total momentum in each direction. Hence, total change of momentum in the direction of motion is $(E/c^2)v$, and since the particle is losing the same momentum expressed as mv, we obtain:

$$m=E/c^2.$$

One cannot deny the simplicity and brevity of the derivation, but does it suffice as a "proof" (by pure thinking), and how general it actually is? In his derivation Einstein relies on:

- law of conservation of momentum – that it holds for light as well as material particles;
- formula for the momentum of light (waves or photons alike) from Maxwell's theory;
- the Lorentz contraction coefficient known from experiments of Fizeau (known to Einstein at the time of the first derivation) and Michelson-Morley experiments (which Einstein was adamant he did not know whilst in process of discoverig STR).

Surely, we could accept this derivation just as given in a thought experiment as a proof in the sense of an *a priori* proof that visible light carries inertia. However, the question remains: what does it take to generalise the deduction to all forms of, first, electromagnetic energy and, then, to all forms of energy.

To generalise the conclusion: that *all electromagnetic energy* (light) has inertia, he assumes that all the different EM-waves differ only by frequency/wavelength and that all the emission or absorption processes are equivalent, in the sense that all the above assumptions/facts hold for any of them. But what does it take to generalise the result to *all energy* has inertia, as he was later to do? General validity of the laws of conservation for all matter-energy, new concept of mass-energy, Lorentz transformations for momentum-energy, Noether's theorems…? These were not all spelled out in the original TE or its versions, as Einstein knew he could do only as much when it came to generalizing his results by using TE as the only tool. However, his equation does hold generally, for all forms of energy, known and yet perhaps to be discovered, and to motivate this we would need to expand our discussion much more to examine the general framework of the relativity theories and the physical and philosophical reasons as to why it should hold, or

why the theories to be discovered should also be expected to be relativistically invariant theories. If we had all these clearly spelled out, we could, I think, be excused, for believing that the argument derived from TE alone suffices to justify the belief in the correctness of the deduction. As Miščević puts it:

> First, note that the alleged minuses of TEs are not really minuses of thought experimenting as such, but rather *deficiencies of available wider frameworks!* Further, if an important thesis is scientifically testable in some reasonable time, then TEs teaching us about it can still be very useful. (2022: 118, my italics.)

So even though I agree with Miščević's overall analysis of generic TEs, I would still point to how truly surprising is not only the result of the light momentum TE, but also the fact that it illuminates aprioristic deductive thinking, albeit in a limited domain of application, something more akin to a Platonist account of TEs rather than a naturalist one along the lines of Miščević's proposal (cf. Brown 1991/2005: especially ch. 4).

6. *Einstein's cocksureness and the reason why scientific and metaphysical TEs could be regarded as special*

Einstein was famous for his open-mindedness when it comes to the potential revision of his, at times even most cherished, beliefs as well as for his honesty in admitting errors of judgment (cf. eg. Janssen 2014: 216). He was also pretty cocksure. When asked by biographer and philosopher Ilse Rosenthal-Schneider about how he received the news from Eddington's solar eclipse expedition of 1919, the results of which proved Einstein's calculations of the bending of light rays from a distant star passing the Sun as predicted by general relativity, he was not exhilarated as expected but laconically retorted:

> 'I knew that the theory is correct. Did you doubt it?' I answered, 'No, of course not. But what would you have said if there had been no confirmation like this?' He replied, 'I would have had to pity our dear God. The theory is correct all the same.' (Rosenthal-Schnieder 1980: 74)

This response could seem nothing short of blasphemy, not to say arrogance. But, of course, it wasn't Einstein's intention to be either. He liked to couch his musings on the nature of physical reality, his philosophical outlook, the meaning of life and other big questions in theological terms, not necessarily adopting any particular theology. He had no interest in being arrogant, especially late in life and after achieving not only the main results of his physics, but also worldwide fame reaching far beyond the community of physicists or scientists in general. So should we take it for granted that Einstein knew when he was right, even before having been given experimental data, that he had some special insight into the nature of things, a direct line to God? Although it is tempting, we should be reminded that Einstein did exclaim in the

past, and on more than one occasion, that he was certain he was in possession of the true theory when he in reality was not. For instance, in 1914, in a letter to Michele Besso, his lifelong friend from student days, Einstein wrote about his perfect satisfaction with the prototype of a general theory of relativity but which lacked the general covariance (the so called *"Entwurf" theory*):

> Now I am completely satisfied and no longer doubt the correctnes of the whole system, whether the observation of the solar eclipse works out or not. The sense [*vernunft*] of the matter is too evident. […] The general theory of invariants functioned only as a hindrance. The direct path proved itself to be the only passable one. (As quoted in Norton 1995: 61–62)

Notice that exactly the opposite was to ultimately show iteslf to be true, namely, that the final theory was to be a covariant theory with certain invariants seen as the crucial part of the whole system, and that the line of thought Einstein was, even against his own will, forced to follow the one of mathematical simplicity and elegance which he will much later praise in his Oxford lecture quoted above, not the direct path of physical insight. Yet, the phrasing is almost identical to the response he gave Rosenthal-Schneider, up to denying the relevance of the solar eclipse results (which are historically the *observatio*, if not *experimentum crucis* for general relativity!). Again, what are we to think of Einstein self-confidence, was it a mere joke? Obviously not to him, as the letter to Besso testifies, where he was in earnest about what he was saying, even if we could doubt the same being true in case of the conversation recorded by Rosenthal-Schneider. The plain matter of the fact is that Einstein was not always sure and could not always be sure about the correctness of his theoretical constructs, and that it would be a mistake to take for granted that he somehow always knew. However, it is also quite evident that Einstein did have a special insight into the nature of things as witnessed by his light momentum TE and so many other similar examples. Einstein was, as said at the beginning of this paper, a master of manipulating thought experiments so as to reveal Nature's secrets – this is what he presumably meant by the *direct path* – and one could not blame him that he preferred this direct, or at least more starightforward, pathway into Nature's hidden realm, not least as it usually served as a kind of shortcut. That sometimes he could not find appropriate shortcut, or that sometimes there was not one, but only the ardous path of abstract mathematics was available, surely cannot be taken against his general outlook. We could say of Einstein as it was said of Benjamin Franklin, and even more trully: *eripuit fulmen coelo sceptrumque tyrannis*. I would also maintain that his cocksuredness was not only a byproduct of his method of thought experimenting and coming to the far reaching conclusions about the nature of things, but that it was a prerequisite for it, as Einstein was first and foremost a theorist and the purest of the pure, but not a stranger to all experimenting (after all he spent some years in the Bern patent office).

It was essential to his method as a theorist to be able to not only do the so called *back-of-the-envelope calculations* but also to try to guess at the solutions before even attempting to solve (or put forward) an equation. In order to develop this kind of method you need cocksuredness as part of your character. I think much here is explicable rationally, but there is a residuum which escapes any rationalistic or naturalistic analysis and is best described by my deliberately chosen words *insight into the nature of things*.

Finally, let me remark on the claim put modestly by Miščević (2022: section 6.4) that he sees scientific and, broadly speaking, philosophical TEs as on a par, and although he can see that scientific TEs usually always have some effect on the discussion at hand (even if disproved by real experiments), he does not see any reason why we should be forced to decide on the comparative value of either based on usefulness only. My response would be, based on the studies of primarily scientific (in particular Einstein's) TEs, but also the metaphysical ones, and comparing them with related *genera* as Miščević espouses throughout his book, that one could potentially try to mount a serious objection to the claim that all the TEs, or rather IETs, stand equal in terms of epistemological value. Namely, behind every scientific (and I would also say metaphysical) TE there must be a general framework which Miščević also mentions in the passage quoted in the previous section of this paper, within which there is a hierarchy of statements, from axioms/hypotheses of highest degree of generality to more specific claims; there is, in Leibnizian jargon, a whole hierarchy of reasons which can justify this or that belief. I am quite doubtful as to the existence of such principles in such areas of philosophy as ethics, politics or philosophy of history. Let us take ethics as an example. If we take ethics heteronomously, then its foundations are outside it, so its grounding principles are not ethical. If, on the other hand, we take it autonomously, we end up with all the intricacies of the problem of the relationship of individual interests as against the interests of the group. And even a theonomously considered ethics has its problems as the main reasons for something happening to us, the grounding principles, are perhaps for ever to stay unclear to us humans (take the Biblical example of Job, who keeps suffering as a righteous man, something that should not be happening according to general morality that comes accross in the *Old Testament*). Does anyone believe, to borrow a phrase of Herman Melville from his *Moby Dick* (ch. 64), that *angels are nothing more than sharks well governed*? Are there universally knowable universal rules of ethics to be found by some thought experimenting such as John Rawls'? And one would be hard pressed indeed to try to find the laws of history or politics. As the course of fate of both individuals and societies is determined by so many factors, it is impossible to know its turns, and the so called *real politics* is usually just bestial, so the only rule is the rule of the jungle.

Acknowledgments

I would like to thank the late Professor Nenad Miščević, my mentor and friend, and Professor James Brown for inviting me to participate in the 47th Philosophy of Science conference in Dubrovnik in April 2022 and contribute to the discussion of Professor Miščević's valuable book. This paper is dedicated to the ever-lasting memory of my Professor.

References

Brown J. R. 1991/2005. *Laboratory of the Mind: Thought Experiments in the Natural Sciences*. London: Routledge.

Einstein A. 1919/1988. "What is the Theory of Relativity?" English translation in A. Einstein 1954. *Ideas and Opinions*. New York: Bonanza Books.

Einstein A. 1934. "On the Method of Theoretical Physics." *Philosophy of Science* 1: 163–169.

Einstein A. 1954/1988. "Letter to Felix Pirani February 2nd 1954." In A. Einstein. *Ideas and Opinions*. New York: Bonanza Books.

Janssen M. 2014. "'No Success Like Failure...': Einstein's Quest for General Relativity, 1907–1920." In M. Janssen and C. Lehner (eds.). *The Cambridge Companion to Einstein*. Cambridge: Cambridge University Press, 167–228.

Miščević N. 2022. *Thought Experiements*. Cham: Springer.

Norton J. D. 1985. "What Was Einstein's Principle of Equivalence?" *Studies in History and Philosophy of Science* 16: 203–246. Reprinted in D. Howard and J. Stachel (eds.). 1989. *Einstein and the History of General Relativity: Einstein Studies* Vol. I. Boston: Birkhäuser, 5–47.

Norton J. D. 1991. "Thought Experiments in Einstein's Work." In T. Horowitz and G. J. Massey (eds.). *Thought Experiments in Science and Philosophy*. Lanham: Rowman and Littlefield Publishers, 129–144.

Norton J. D. 1995. "Eliminative Induction as a Method of Discovery: How Einstein Discovered General Relativity." In J. Leplin (ed.). *The Creation of Ideas in Physics: Studies for a Methodology of Theory Construction*. Dordrecht: Kluwer, 29–69.

Norton J. D. 2014. "Einstein's Special Theory of Relativity and the Problems in the Electrodynamics of Moving Bodies That Led Him to It." In M. Janssen and C. Lehner (eds.). *The Cambridge Companion to Einstein*. Cambridge: Cambridge University Press, 72–103.

Norton J. D. 2021. "Author's Responses." *Studies in History and Philosophy of Science* 85: 114–126.

Rosenthal-Schneider I. 1980. *Reality and Scientific Truth*. Detroit: Wayne State University Press.

Russell B. 1937/1992. *A Critical Exposition of the Philosophy of Leibniz with an Appendix of Leading Passages*, 3rd ed. London and New York: Routledge.

Stachel J. 1989. "The Rigidly Rotating Disk as the 'Missing Link' in the History of General Relativity." In D. Howard and J. Stachel (eds.). *Einstein and the History of General Relativity: Einstein Studies* Vol. I. Boston: Birkhäuser, 48–62.

Synge J. L. 1960. *Relativity: The General Theory*. Amsterdam: North-Holland.