Apriorism, Psychologism, and Conceptualism about Thought Experiments

Abstract:
Epistemological optimists about thought experiments hold that imagination could be under certain conditions source of epistemic justification. Their claim is usually based on one of three dominant conceptions about epistemic value of thought experiments. Apriorism states that imagination may serve as unique a priori source of new synthetic knowledge about the actual world. I argue against this view and show that apriorism is either too weak, or too strong or too vague. Psychologism is viable, yet not fully clear conception about new meta-knowledge obtained by thought experimenting. I compare some interpretations of this position and present reasons for favorizing one of them. Conceptualism considers thought experiments as instruments for cleaning our conceptual systems. I argue that this position is in fact not about epistemic value of thought experiments, but about one specific usage of experimental result.

Keywords:
thought experiment, epistemology, apriorism, psychologism, conceptualism

I define thought experiment as *a set of instruction that tells us what to imagine with aim to get to know something* (Picha 2011, 22): the contemplation of thought experiment leads to an explicit acceptance of proposition. The principal epistemological question is whether the belief obtained this way may be taken as justified. Do we have a reason to take this belief as true?

Imagination, which is the core of thought experimentation, is a confusing process influenced by individual history, dispositions, opinions, temperament and preferences. Generalizations are extremely difficult because of the subjective nature of imagination. However, despite all the doubts about thought experiments we still use them, we pay attention to them and sometimes we are even convinced by them. Researchers quite dramatically diverge in opinion whether we should let thought experiments convince us. On one hand we find *optimists* who understand thought experiments as a solid source of justification. There is nothing wrong with the use of thought experiments in the search for truth, and if we are sufficiently careful experimenters, we can obtain new knowledge in this way. On the other hand, there are *skeptics* who deny any epistemic value of thought experiments: Imagination is misleading and cannot serve as a source of justification. Thought experiments can be used only as didactic tools; their epistemic value is zero.

The following paper introduces three optimistic positions: *apriorism* which understands thought experiments as unique source of knowledge about actual world; *psychologism*, according to which thought experiments show something new about our beliefs; and *conceptualism* that attributes to thought experiments a significant role in the elaboration of our conceptual equipment.

Brown's apriorism

Let's start with a non-empirical conception. Its proponent John Brown argues that some thought experiments can convey information about our physical world which has not been obtained by empirical way or derived in the logical sense from information already known. In this sense, some thought experiments are the source of synthetic *a priori* knowledge (Brown 1991, 76). Brown's concept of thought experiments is *en bloc* Platonist. It is a combination of epistemological claims about the way thought experiments produce beliefs (apriorism), and metaphysical claims about the status of objects and relationships related to these beliefs (realism of ideas). Platonism assumes a special non-sensorial epistemic channel leading to objective, independent, abstract objects, which provides new synthetic information about the world. Some thought experiments therefore do not merely transform beliefs obtained through epistemically conventional channels; they are rather "armchair drills" which can excavate a new way to universals founding natural laws. Thought experiments are reliable because they provide information from ideal domain.

What are Brown's arguments to justify his conception? At first, he tries to diminish its extraordinary character by showing that Platonism in mathematics is regarded as legitimate and acceptable. This argument is certainly not strong enough to fully support Brown's statement about thought experiments; however, his objective is mainly to reduce reader's resistance to his conception, because of the authority of mathematics and mathematicians. His second argument is more interesting and requires a larger explanation. Basically, it means to identify certain group of thought experiments whose functioning cannot be explained by empiricism.

Taxonomy

There are many ways to sort out thought experiments; Brown proposes taxonomy based on three criteria. First, he considers, in accordance with Popper (1959, 521), destructive and constructive thought experiments. Destructive experiments mean to criticize a theory, while constructive experiments serve to support it. – Secondly, constructive thought experiments can be divided according to the character of imaginary situation. How can we understand that? Brown describes the principle of thought experimenting as the "establishing a phenomenon", i.e. obtaining the very result of experiment. Brown then distinguishes between thought experiments with a surprising result (i.e. established phenomenon) that requires further explanation, and thought experiments with a non-problematic result that does not require any further explanation. – Thirdly, Brown asks whether the criticized or supported theory is formulated before the experimentation, or if the explicit formulation of the theory is an outcome of thought experimenting. So there are three criteria – relation to a theory, problematic character of result, status of a theory – used by Brown to identify destructive thought experiments, as well as three types of constructive thought experiments: mediating, conjectural and direct.

	Theory before experiment	*Theory after experiment*
Problematic result		conjectural
Non-problematic result	mediating	direct

Brown's typology of constructive thought experiments

Mediating thought experiments are basically didactical tools. Their aim is to make difficult theories accessible by the means of a non-problematic phenomenon. Brown shows that mediating thought experiments can function as illustrations of non-intuitive outputs or as diagrams which help to understand formal aspects of theories, ex. Maxwell's Demon. – *Conjectural* thought experiments establish a problematic phenomenon which requires a new theory. Conjectural experiments are easier to refuse by criticizing that the phenomenon established by the experiment would not actually occur, i.e. that the experimental output is false. – *Direct* thought experiments combine non-problematic character of mediating thought experiment with the epistemic value of conjectural experiments. They are based on establishing of a generally acceptable phenomenon which leads to the formulation of a new theory capable to explain the phenomenon.[1]

From a perspective of epistemic value, the *platonic* thought experiments are the most important category of Brown's taxonomy. They are scenarios with both destructive and constructive role: they criticize one theory to replace it by another one. Platonic thought experiments combine the ability to defeat a theory with the constructive qualities of direct thought experiments and their non-problematic phenomenon. A platonic thought experiment – a typical example is Galileo's Pisa Experiment – simply refuse one theory to replace it with a new, better one.

What is the relation between presented taxonomy and evaluation of epistemic value of thought experiments? We have to turn to platonic experiments used by Brown to support his apriorism – they produce, according to Brown, a new synthetic knowledge *a priori*. Brown supports his argument by a reconstruction of Pisa Experiment to show more precisely when we acquire an *a priori* knowledge about natural laws of actual world. Contemplating Pisa Experiment is supposed to produce knowledge that cannot be deduced from already known information.[2]

Objections against apriorism

Why have thought experiments an epistemic value? It seems that the answer of the aprioristic conception is brutally simple: because they are *a priori*. I argue that such answer is insufficient and based on a trick. Apriorism alone is a conception dealing with the *acquisition* of knowledge only, not with the *justification* of knowledge. The trick lies in the fact that traditional epistemological conceptions treat *a priori* beliefs as analytical truths. If

[1] For an example of conjectural and direct thought experiment see Brown (1991, 3-10).

[2] For the detailed critique of Brown's reconstruction of Pisa Experiment see Picha (2011, 91-124).

something is *a priori*, then it is necessary and true. In such conceptions *a priori* is not only a way we acquire belief, but also a reason why the belief is true. However, in the context of Brown's theory, we have an apriorism that does not deal with analytical truths. It is only a description of acquisition of beliefs – those beliefs which cannot be explained by an *a posteriori* way – but there is nothing that justifies truth of these beliefs. Apriorism is rather a psychological conception than an epistemic one.

Generally speaking, there are three types of objections against apriorism. First, they question whether the epistemic value of thought experiments must be explained by the acquisition of *new* data. Secondly, they consider the guarantee of truth of *a priori* beliefs. In the particular case of Brown's version of apriorism, there are doubts whether his Platonist assumptions is in accordance with the principle of simplest explanation. The objection is not exceptional in the context of thought experiments and appears frequently in general discussion on Platonism. Third objection deals with the possibility of *mistake* in the context of apriorism. Nobody doubts that thought experiments can be misleading and produce beliefs which are not true. Besides the explanation of reliability of thought experiments, we have to explain their fallibility as well. Brown (1991, 92-93) considers the third objection as possibly the most important in the philosophy of science; on the other hand he marginalizes it by comparison with empirical fallibility. He says that *a priori* fallibility is not a greater problem than the fallibility of our senses. Whether we agree with Brown's comment or not, it is an acknowledgement of weakness that does not appear in other conceptions of epistemic value of thought experiments. They do not postulate a specific epistemic channel and thus do not need to formulate their own theory of mistake. The objection can be even stronger – apriorism is not only too complicated, it is also incomplete. It considers only a limited amount of thought experiments (platonic) and does not give account of the rest. We should probably admit that mediating experiments are put aside because of the absence of epistemic value; let us be generous and assume that epistemic value of direct experiments is determined the same way as the platonic experiments. But what are we to think about conjectural thought experiments? Should we expect that they lead to new data by the means of problematic phenomenon? Shall we consider as epistemically valuable also the experiments where our results diverge or change in time? If it is true, then we certainly need a theory of fallibility to be able to correct our evaluation.

Problems of problematic nature

I have to express as well a critical objection on Brown's taxonomy, even though it is just a tool for a better presentation of apriorism. In long run, any taxonomy needs precisely defined criteria. Brown's taxonomy has the weakness that lies in the vague character of criteria, especially the problematicity of phenomenon. Brown explains his conception at two places. First, he considers direct thought experiments are based on non-problematic phenomena, unlike conjectural experiments (Brown 1991, 41). Secondly, he characterizes conjectural thought experiments as the target of objections that doubt phenomena (Brown

1991, 40). Following the formulations I understand that the non-problematic phenomenon established by the experiment is in accordance with our intuitions, which leads to the acceptance of the phenomenon. A problematic phenomenon is established by the experiments but is not in accordance with our intuitions.

Troubles with this conception of problematicity lie in the way we form our relevant intuitions about hypothetical phenomena. More precisely, I find problematicity of phenomenon as an individually, historically and culturally based evaluation which can be hardly generalized; ex. a problematic phenomenon in Renaissance would not be considered as problematic five centuries later (for instance, seven billion people on Earth, instant communication in long distances, microsurgeries etc.) The same objection applies in the case of experimenters who differ culturally, socially or by acquired education. Let me be clear, I do not hold that it is impossible to obtain generalizations about our intuitions which could be used to categorize experiments. It is certainly possible to conduct a research of opinions on problematicity of hypothetical phenomenon, and divide thought experiments in two groups: experiments with rather problematic phenomena and experiments with rather intuitively acceptable phenomena. I only think that the proposed categorization of thought experiment is relative to various externalities and may change. A conjectural experiment for us can very well be a direct experiment for our children.

Let me add a final remark. Maybe we should understand non-problematicity of phenomenon in a different meaning. Maybe Brown wants to differentiate phenomena by the acceptance or non-acceptance of theory which could adequately explain the phenomenon. Problematic phenomenon would be established and intuitive, nevertheless still unexplained by current theory. Problematicity would be defined in relation to accepted theory, not intuitions. If Brown really means this conception of problematicity, then his division between conjectural and direct experiments does not make good sense. In the two categories, the theory is built after the thought experiment and phenomena in both categories should thus be taken as problematic. However, Brown explicitly states that direct thought experiments establish non-problematic phenomena. This interpretation of problematicity cannot be plausible then.

Apriorism is a daring conception that is not afraid of radical epistemological and metaphysical additives. This could be applied to show that the debate about epistemic value of thought experiments is far from being marginal. If a reasonable person is willing to accept robust epistemological and metaphysical premises to explain thought experiments, it is probably a topic which deserves our attention. Other optimistic conceptions do not share this radicalism. They try to explain epistemic value of thought experiment in accordance with the assumption that synthetic knowledge about world can be obtained *a posteriori* only. In this respect, the most influential conception is the theory by Ernst Mach.

Mach's psychologism

According to Mach (1960, 27-28), thought experiments are credible because they are based on usual and generally accepted source of justification – perception. In contrast to apriorism, the aim is not to create a new epistemic channel, but to use an old one in a new way. Mach assumes the existence of "instinctive knowledge" – perceptively, even though unconsciously justified beliefs[3]. Thought experiments serve to pull instinctive knowledge out of the dark parts of our mind to the bright place in our attention. While performing thought experiment we realize that we already have certain primary, perceptually justified belief. From psychological point of view, imagination is the source of new belief – meta-belief that does not concern the world but our beliefs themselves. Now we have to ask an epistemological question: primary belief is justified perceptually, what is the justification of the meta-belief? The answer is straightforward: the source of justification of meta-belief is the same as in the case of beliefs about our other mental states – introspection. From this point of view, thought experiments are the instruments of introspection which make our hidden non-reflected beliefs explicit. We can see it analogical to the process of filling a room with fog to be able to observe the traces of microparticles. Fog is the analog of imaginary situation which creates environment suitable for explicit disclosure of objects or beliefs that have not been observed before.

Mach is not specific about his conception, so it is not possible to have a precise idea about the character of mental processes that could lead to disclosure of non-reflected beliefs. Sorensen (1992, 88) describes several ways how to "improve the epistemic status of thinker without the addition of new information". Three processes are worth to be mentioned: remembering, transformation and rearrangement.

Remembering means to recall old information thanks to imagination; it is a change of dispositional belief into occurrent one. Sorensen connects this conception with Mach's conception of thought experiment, but I consider the connection as superficial. It is generally acknowledged that in order to talk about remembering, a memory must have been occurrent belief before. For example, in order to recall now the lunch you had yesterday, you would have to be aware of what you were eating during your lunch. Mach's conception of explication of beliefs does not count with such conditions. Thanks to thought experiments we can now become aware of belief which was never occurrent.

Is not this conception of remembering past occurrent beliefs only too narrow? Let us count as remembering also the recalling of perceptual information we were never aware of before. For example, in a state of hypnosis we are able to remember details which have never been the content of subject's occurrent beliefs. Even such broadened notion of remembering does

[3] Externalist has no problem considering instinctive knowledge as knowledge. Internalist prefers to talk about justified true proto-belief.

not correspond to Mach's conception. According to him, thought experiments can explicate even a belief *not obtained by perception*. This claim obviously raises questions about Mach's epistemology: How is existence of non-perceptual beliefs compatible with his declared empiricism? Mach surprisingly combines empiricism with innatism. Some of our beliefs are innate and thought experiments are able to disclose them, ex. fear of heights, water and foreigners. We have not obtained these beliefs by perception; nevertheless, we can make them explicit by thought experimenting. Those explicit beliefs are not memories of non-conscious *perception*. At the same time, Mach holds that even the innate beliefs are somehow perceptual, because one of our ancestors obtained them by sensory organs. Thought experiments thus can explicate beliefs which are perceptual from the point of view of subject or his ancestors. However, the explication is not a recalling, because we cannot possess someone else's memories, not even memories of our ancestor.

Transformation is a change of order of old data to the form which is easier to handle. It is a change of the relation of belief to surrounding beliefs. Let us have a look at a hidden non-reflected belief stating that sea is mortally dangerous. The belief is in relation to other, reflected beliefs, ex. that there are sharks, medusas and toxins in the sea, that it is impossible to breathe in the water and there is a high pressure. Accumulation of these reflected beliefs in imaginary scenario leads to the discovery of their inferential connections to the non-reflected belief about dangerous sea. I imagine swimming above a mass of water, the view of my legs seen from bellow and I realize that such image is not a pleasant one. Based on my imagination, I realize that I suffer from belief about mortal danger hidden in the sea. Thought experiment turned attention to inferential links between my beliefs and made explicit the hidden one.

Rearrangement is a change of ways to treat and keep old data; it is a transformation of coding the information. Let us suppose you obtain a following set of beliefs while reading a fairy tale: princess is blond, thin, tall, pretty, with a lovely voice, a Mediterranean type, probably suffering from anemia and romantic ideas and waiting for her prince. Then you will try to use your imagination and create her image before your mind's eye. You will realize that it is not possible to join consistently beliefs about her being blond, tall, thin and Mediterranean type. The resulting schism can be described by the fail of attempt to process propositional information by the center of visual information. It resulted in the explication of belief that princess cannot be tall thin blond and Mediterranean type at the same time. Even if the rearrangement can vary, in practice it is always visualization, because imagination is almost always powered by images. Of course, there are talented and experienced experimenters able to imagine auditory, olfactory, sensational or tactile sensations clearly enough to be considered as a basis for thought experiment with tones, smells, meals or textile, but generally speaking, human beings rely primarily on visual impressions. In comparison with other senses, sight is highly distinctive source of information. A

transformation model explains thought experiments as visualizations of information obtained by epistemically credible ways, which enables to acquire new beliefs on their basis.

I believe only the transformation model is in accordance with Mach's conception. Why does not the rearrangement model correspond to Mach's view? Simply because Mach supposes an existence of primary belief, that is waiting to be discovered. In the dark part of our mind, there is a completely made belief; we only need to bring it to the light. The rearrangement, however, is not the same – old information is newly treated and lead to the formation of brand new beliefs. In the rearrangement model, imagination helps to acquire new primary belief, not to obtain meta-beliefs about existing primary belief. It is not, however, important which model of transformation suits best to Mach and his conception; the only important thing is that there are some plausible *empirical* explanations of epistemic value of thought experiments.

Mach's conception has been of great influence. Mach proposed a positive answer to the question whether thought experimenting may be source of justification. He parts from the supposition that thought experiments are not epistemologically unique instruments. They are just common ways of thinking utilizing the facts that our minds are not transparent and that we can introspect. In this sense Mach can be seen as initiator and defender of "ordinary approach" to thought experiments, i.e. the opinion that epistemic value of thought experiments can be fully explained by our ordinary cognitive abilities[4]. Eventual doubt about epistemic value of thought experiments is not *sui generis*; it is doubt about the credibility of introspection.

Kuhn's conceptualism

The conceptions mentioned above differ in their opinion about the ways in which thought experiments produce new synthetic beliefs. According to Brown, we learn about the actual world by learning about the ideal world; according to Mach, we learn about the actual world by learning about our beliefs. Kuhn believes that we learn about the actual world by learning about our concepts. Kuhn's approach is divided into two parts: the first one deals with the way thought experiments produce new beliefs, the second one shows how new beliefs relate to the actual world (Kuhn 1977).

Let us consider the first part. Contradiction is the central notion in Kuhn's theory. Contradiction usually describes a proposition that is always false, or a pair of propositions whose conjunction is always false, ex. 'x is F' and 'x is not F'. Contradiction is thus property of proposition, or more precisely, of systems of propositions. Sometimes we talk about contradictory concept in the sense of *contradictio in adjecto* when sub-concepts of a contradictory concept identify properties impossible to be instantiated together at the same

[4] Perspective defended and popularized recently by Timothy Williamson (2004). Similar approach is developed in van Inwagen (1988).

time, ex. ‘round square’. Kuhn thinks that there are many concepts in use that contain hidden contradictions. He uses Piaget’s real experiment: when little children were asked to show the fastest car out of a group of cars, the children repeatedly chose the car which reached the goal as the first one and did not pay attention to the starting point of the cars. The children did not consider the distance, only the time needed to reach the end. At the same time, the children decided according to other, phenomenal criteria. They describe as faster the car that ‘seemed to move faster’ than the other one[5].

Kuhn interprets the experiment by stating that children have a contradictory concept of ‘faster’; and thought experiments are able to discover the hidden conceptual contradictions. Let a child imagine a situation in which a car is moving faster according to phenomenal criteria. At the same time, the car will reach the end later than the other car that either starter first or was closer to the end at the start than the first one. The child will realize that under such conditions the car would be faster and slower than the other car at the same time, which is absurd. The experimenting child will obtain a new belief about the contradictory concept leading to paradox.

Let us consider the second part: How does the new belief about the contradictory concept concerns actual world? The contradictory concept ‘faster’ differs from the contradictory concept ‘round square’ in the sense that the children’s concept ‘faster’ is not a necessary contradiction. The notion ‘round square’ is internally incompatible, while the children’s notion ‘faster’ is not – it is possible to imagine a world where all objects move with the same speed. In such world, children’s concept ‘faster’ will not be contradictory because phenomenal criteria will be useless. The concept ‘faster’ will then simply correspond to the concept ‘to finish first’. Kuhn proposes the following: Children’s concept ‘faster’ is not necessary contradiction, but only contingent one. The applicability of the concept depends on specific natural laws of the possible world. By discovering contingent contradictory nature of the concept we realize we do not live in a world with certain natural laws – which is new synthetic knowledge. Thought experiment is then a source of knowledge about the actual world; it is informative, because by contemplation we realize what does not apply in our world.

I find Kuhn’s conception of thought experiments limited, though somehow easier to be accepted than some other conceptions. Let me explain my criticism; why is his conception limited? Kuhn describes only the argumentative *use* of imaginary scenarios. He does not say how and where we obtain our beliefs, or why we should trust such beliefs. He just tells us what to do with such belief. To be specific, Kuhn does not offer an explanation why the belief that a car can move phenomenally faster and finish second is justified. – Moreover, Kuhn takes in consideration only thought experiments having the form of *reductio ad*

[5] Kuhn (1977, 244) mentions he will use the term ‘perceptual blurriness’.

absurdum. Thought experiments show a situation which enables us to see incompatible inferential results of an accepted belief, ex. children's definition of the concept 'faster'. The following structure can be found in some examples only. There are numerous thought experiments that cannot be reasonably reconstructed as *reductio*, ex. Leibniz's Mill, Maxwell's Demon etc. Kuhn's conception is not only limited in the sense that it concerns broad application of thought experiments; it also deals with just one type of several possible applications.

At the same time, Kuhn's conception is somehow more acceptable. Thought experiments are shown as reasonable and perspective tools of scientific research. A great advantage of *reductio ad absurdum* is the fact that it starts with the statements of the opponent. The premise is – at least provisory – by both sides considered as accepted starting point. Furthermore, Kuhn's support for the claim that thought experiments concern actual world is almost trivial. Why? As I understand Kuhn's position, he explains thought experimenting as evaluation of two arguments. In the first one, we have to imagine certain situation and thus derive the metaphysical *possibility* of such situation. The situation is then used as a counterexample to opponent's statement. To be specific again, we imagine 'phenomenally faster' car finishing second which serves as a counterexample to the children definition of 'faster'. In the second argument, we imagine a possible world *w'* where the experimental situation cannot occur – and then we derive the contingency of such situation. That is the way, according to Kuhn, thought experiments bring information about principles of our world. We learn about what does not apply in our world; we realize that *w'* is not our world. Do we really find out anything new about our world?

The example of car offers two possible – because conceivable – situations. In the first one, the cars can move with different speed; in the second one, all cars have to move with the same speed. It is the first situation that deals with our world and we know *even before* the experiment that objects move with various speed in our world. Considering experiment we just explicitly realize that we do not live in a world where the principle of constant speed applies, nothing more. The following example illustrates the triviality of Kuhn's position. We know that the speed of light in vacuum in our world equals to *c*. We can imagine a world with a different speed of light in vacuum. Thanks to that, we realize we do not live in a world where the speed of light in vacuum is different than *c*. Analogically, we know that the objects move with various speed in our world. We can imagine a world where it would be different. Thus, we realize that the principle of constant speed does not apply in our world.

The thought experiment with cars is an instrument that teaches us two things: first, imagined situation with cars is metaphysically possible; secondly, the imagined situation with cars is nomologically impossible. Only the first information is obtained by thought experiment, i.e. by the use of imagination. The second one is obtained by a banal *a priori* deduction from the knowledge of the actual world. Thought experiment is the source of

synthetic knowledge just because new conclusions about metaphysical status can be combined with our old knowledge of the actual world – and we realize what could apply but actually does not.

Kuhn's conception is in fact not an optimistic conception of epistemic value of thought experiments. It is rather an optimistic conception of usefulness of thought experiments in science. Kuhn does not say how we reach the results of thought experiments; he says how the results are used in scientific argumentation and why they are sometimes useful even for scientists. Unlike apriorism or psychologism, conceptualism is only a partial comment on the function of thought experiments, not a theory about their essence or epistemic value in general.

Marek Picha

Masaryk University
Faculty of Arts
Brno, Czech Republic
picha@phil.muni.cz

References:

Brown, J., 1991: *The Laboratory of the Mind: Thought Experiments in the Natural Sciences.* London: Routledge.

Kuhn, T., 1977: *The Essential Tension.* Chicago: University of Chicago Press.

Mach, E., 1960: *The Science of Mechanics: A Critical and Historical Account of its Development.* La Salle: Open Court.

Mach, E., 1976: *Knowledge and Error: Sketches on the Psychology of Enquiry.* Dordrecht: Reidel.

Picha, M., 2011: *Kdyby Chyby: Epistemologie Myšlenkových Experimentů.* Olomouc: Nakladatelství Olomouc.

Popper, K., 1959: *The Logic of Scientific Discovery.* London: Hutchinson.

Sorensen, R., 1992: *Thought Experiments.* New York: Oxford University Press.

Inwagen, P. van, 1988: Modal Epistemology. *Philosophical Studies*, 92, 67-84.

Williamson, T., 2004: Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking. *Proceedings of the Aristotelian Society*, CV, 1-23.