Croatian Journal of Philosophy
Vol. XXIV, No. 71, 2024
https://doi.org/10.52685/cjp.24.71.1
Received: September 19, 2022
Accepted: June 26, 2024

Miščević: Mental Models and More

JAMES ROBERT BROWN
University of Toronto, Toronto, Canada

This is a review discussion of Nenad Miščević's stimulating new book, Thought Experiments (2022). His mental models account is of great importance in the various current debates about the nature of thought experiments. I discuss some of the pros and cons of his account.

Keywords: Thought experiments; mental models; intuitions.

Thought experiments (TEs) are remarkable devices for producing knowledge. Physics and philosophy are full of them, and it would be hard to imagine either discipline progressing as they have without a heavy dose of the kind of imaginative thinking produced by TEs. Galileo's ship, Einstein's elevator, Schrödinger's cat, and a great many more have played a central role in the development of physics. Plato's cave, Leibniz's mill, Putnam's twin earth, the trolley problem have similarly enriched and shaped the course of philosophy. In his new book Nenad Miščević offers a justification that I think we can all endorse. "Thought experiments are indispensable. Philosophy does not use a laboratory to test its theories; the only experiments available here are those in thought. TEs play in philosophy the crucial role that laboratory experiments play in science. Philosophers are vitally interested in *connections between our spontaneous understanding of important items, like meaning and content of our thoughts, and the results of science*" (2022: 87).

Many questions arise. How do TEs work? What are the different kinds? Why do some disciplines have a lot of TEs while others have few or none? The central question is this: How is it possible to learn something new about reality merely by thinking?

Nenad Miščević has an answer: *mental models*. His account can be found in various of his papers and now in his stimulating new book, *Thought Experiments* (Miščević 2022). Mental models, he claims, can address all (or most) issues concerning TEs. He introduced this ap-

proach in a talk about 30 years ago (subsequently published as Miščević 1992, simultaneously with, but independently from Nancy Nercessian (1993)). I fondly remember the occasion. It was in Dubrovnik in the siege during the violent breakup of Yugoslavia. One could come into the city or leave only by a daily ferry from Rijeka. The Inter-University Centre where the annual philosophy of science conference was held had been bombed and was largely destroyed. So, we met in temporary surroundings elsewhere in the old city. At night we heard machine gun fire. Snipers in the hills helped to focus the mind.

Nenad's mental models account is extremely plausible. As the term suggests, we form a model in our heads then read off the details that are consequences of the model. One of the strongest pieces of evidence for this account comes from our ability to make inference almost instantaneously. Imagine a turtle on a log. A fish swims under the log. Is the fish under the turtle? We immediately say yes, because we can see it in our mental model. A rival account of thinking would have us make inferences (deductive or inductive) from the given premisses. The trouble is that it takes several slow steps to get to the conclusion that the fish is under the turtle. This makes the mental model account much better at explaining how we actually reason in a wide variety of cases. And it makes a great deal of thought experimenting easy to understand as simply being instances of mental modelling.

At one point Nenad remarks, "It is Kant whose account of 'construction in intuition' comes closest to the mental model view" (2022: 61). This might need some explanation, since Nenad is a naturalist and a liberal empiricist, so there could be some tension. But this is a point I will not pursue. Instead, I will note the contrast with my own view. I think that (some) intuitions should be understood as producing genuine new knowledge. This is not a construction in imagination, nor an examination of our concepts, but rather a kind of perception of something existing independently from us. Such an account is anathema to empiricists and naturalists. Serious intuitions involve a kind of intellectual grasp, seeing with the mind's eye.

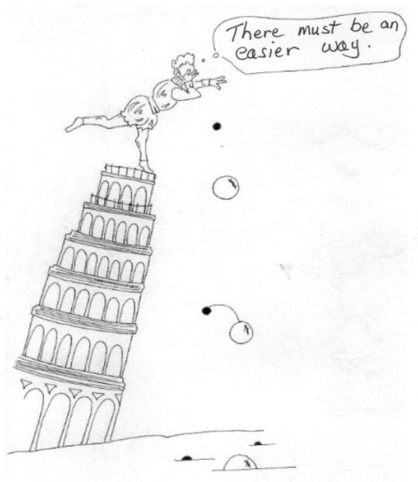
Nenad argues that: "the mental modelling theory and the 'voice-of-competence' proposal can account for most, perhaps even all, puzzling phenomena tied to thought experiments and intuitions" (2022: vii). Evolution comes to the rescue: "The evolutionary, adaptationist hypothesis offers a hope that a part of our primitive intuitional knowledge does reflect the deep make up of our environment, and thus, in spite of its fallibility, carry information about real and philosophically important properties of some states of affairs in the world" (2022: 68). Nenad also says, "Whereas Brown thinks that intuition capacity is a basic capacity, I prefer to think of it is a derived capacity that employs various basic capacities, prominently reasoning and quasi-perceptual imagination in the off-line fashion" (2022: 73). Moreover, he adds, "Intuitions should be studied as any other sources of cognition; one should search

for already known capacities and try to account for intuitions starting from them, instead of ad hoc postulating new capacities" (2022: 74).

Of course, it is difficult to disagree. As a general rule we should not introduce anything new, including new cognitive mechanisms, when we can account for everything with the equipment we already have. Here is a simple example. Those who fish sometimes marvel at the ability of some who seems to know where the fish are. We might say they have great fishing intuitions. I have no such intuitions, nor had my father. It turns out there is a simple empirical explanation for why some people do so well. They have tacit empirical knowledge of the situation. A fast flowing river will create eddies, pools of slow moving water, say, around a large rock. A trout will lurk in such a region because it requires relatively little energy to stay in place. If it stays near the seam of the two regions, the fast water will be a moving buffet, bring food for the hungry fish. None of this might be consciously noticed but it is all empirically absorbed by the alert fisher. Most of our unexplained intuitions will have an empirical source like this. Most – but not all.

I think there are cases, albeit quite rare, where we would be very hard pressed to give a naturalistic account of our intuitions. I will give two examples of this, one from physics and the other from mathematics. The first is obviously a thought experiment; the second is next of kin.

First, a word of explanation. Nenad has introduced useful terminology to cover this. An IET is an imaginative exercise in thought. It covers thought experiments and more, and would include the mathematics example I am about to present. I resist the desire to define thought experiment; I prefer a characterization that sets rough boundaries but does not try to make them precise. A definition can come at the end of inquiry. This is how we treat all sorts of important concepts. Religion and democracy, for instance, are not precisely defined, yet we can rationally discuss them. As for thought experiments, I only want to insist that they be performed in the mind and have an experiential character. We might ask about what the tides would be like, if there were 25 moons instead of one. We cannot "see" the answer; we would need to calculate. So, I would not call that a thought experiment, though others often do. On the other hand, some visual reasoning in mathematics might not be a thought experiment, but it is next of kin. Nenad's term, IET, captures this nicely. Now to my two examples of intuitions that produce genuinely new results.



Galileo first noted that Aristotle and common sense claim that a heavy object such as a cannon ball falls faster than a lighter object such as a musket ball (H > L). From this, it follows that combined cannon and musket balls would fall faster than the cannon ball alone. (H+L > H)

However, the lighter musket ball would tend to slow down the heavier cannon ball with the result that the cannon ball alone would fall faster than the combined object (H > H+L). Thus, we have a contradiction. Aristotle and common sense must be wrong. Galileo was able to resolve the situation by simply having all objects fall at the same speed (H = L = H+L). In other words, all bodies fall at the same rate, regardless of their weight.

This is a truly remarkable result. It is certainly a prime candidate for *a priori* knowledge. Why? There are unquestionably empirical concepts involved, such as weight and falling. But experience did not give us the result; that took the thought experiment. In fact, there was no new experience that moved us from Aristotle's to Galileo's view of falling bodies. The result is not derived from previous experience. Nor is it any kind of logical truth. After all, objects could fall at different rates based on their colour. Those who recall the rise and fall of the fifth force will remember the main claim that different rates of fall would depend on chemical composition. In any case, thanks to this example it can be plausibly claimed that we have *a priori* knowledge of nature. This is something no empiricist or naturalist can entertain.

My second example is from elementary number theory. What is the sum of the first n numbers? A theorem answers this question. The standard proof of this theorem is by mathematical induction, a technique that everyone takes to be a legitimate proof. A diagram is generally considered illegitimate as evidence. Of course, a picture can be pedagogically useful and perhaps helpful in suggesting a legitimate proof, but it is not thought to be acceptably rigorous.

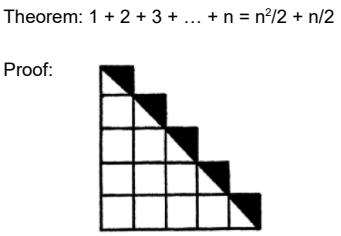


Figure 2. *Picture proof of a theorem.*

Spend a moment on the picture to see how it works. If you need a hint, here it is: Starting from the top add the squares, 1 + 2 + 3 + 4 + 5. Imagine this is a 5 x 5 square. Cut it in half with a diagonal. This represents $n^2/2$. Now restore the half squares (blacked out) that were removed by the diagonal. This represents $n/2$.

After studying the example, you should be persuaded of two things. First, the picture proof is just as rigorous as a proof by mathematical induction. And second, thanks to the first point, intuition is essential. This will be obvious when you realize that the proof holds for every number n, all infinitely many of them, even though the actual diagram is only for the number 5. Needless to say, there are different kinds of intuition, most are compatible with empiricism. Many of Nenad's uses of the term involve cases such as Putnam's twin earth. Here intuition means something like common sense judgement, which is based on empirical experience. As I mentioned earlier, I have no quarrel with these uses and quite agree on their empirical respectability. It is the rare kind that are not empirically respectable that I claim exist. The picture proof and the Galileo case are examples.

I take the Galileo thought experiment and the picture proof of the number theory theorem to demonstrate the existence of genuine knowledge-producing intuitions. I call them platonic intuitions. Such intuitions are not at work in every thought experiment, only a few. We reason about other cases in a variety of ways, as I acknowledged when asserting my pluralism about TEs. Some of these use mental models, just as Nenad claims. In fact, there are a large number of things on which we agree. I should mention some of these, as they are important. The first – and I want to stress this – is that I like Nenad's mental models account very much. It is probably the most popular account of TEs, and for good reason. My own view is often misunderstood, since I embrace intuitions and a generally platonistic outlook. In fact, to repeat again, I am a pluralist about thought experiments. I think Nenad is right about lots of them. I think John Norton, whose view is at the other end of the empiricist-rationalist spectrum from mine, is often right, too. Real experiments work in lots of very different ways. It should come as no surprise that the same is true of thought experiments.

One thing that Nenad took up that is otherwise little discussed is the difference between thought experiments in philosophy and in the sciences. We agree that in some broad sense they are the same kind of thing. Of course, philosophers of science talk about thought experiments in the sciences while regular philosophers focus on thought experiments in ethics, language, mind, and so on. But this is not what I have in mind. There is often a difference in methodological approach. Nenad put his finger on it: "One issue that has been prominent in the discussion is the contrast between 'extroversion' and 'introversion': are intuitions concerned with their external objects, the domain of items and facts, or with our concepts? Is Galileo investigating the falling bodies, or of the concept of the falling body? My sympathies are with external reference. 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: 25). This is a hugely important point and I wholly agree with Nenad. Of course, it is important to know how language and our various concepts work, but ultimately, we are concerned with how the mind-independent world works.

Incidentally, I think every philosopher of science would also agree. This is one of the obstacles to overcome in finally getting something like a unified account of TEs in philosophy and the sciences. Like Nenad, I think that thought experiments are the same in both disciplines, but when some are focusing on *things* and others are talking about *concepts of things*, it can be difficult.

Nenad answers a question I have often raised hoping for an answer. Why are some disciplines more likely to use TEs than others? In particular, why does chemistry have so few, possibly none? Nenad puts it this way and provides an answer: "Why don't we normally have very reliable intuitions about chemistry? A natural answer is that chemical knowledge is not part of our folk theories, and that chemical reactions are not accessible to us to the degree physical reactions are. Therefore, there are no relevant assumptions that a thought experimenter might use. The [mental] models view offers a direction for explaining the phenomenon; I wonder whether the Platonist has anything comparable" (2022: 62).

No, probably not. Nenad's explanation sounds plausible to me initially. But I hesitate to embrace it because it skirts close to a view of TEs held by Daniel Dennett. Dennett has long been a critic of TEs for several reasons. One of these is his claim that TEs rest on folk science. We should expect the world to be very different from our folk conceptions, he says, and therefore, we should really give them no heed at all. We face a two-part problem: first, according to Nenad, we need folk concepts, which we don't have in chemistry, then, according to Dennett, we should reject folk concepts as fundamentally misguided. Consequently, if we need folk concepts but they are misguided in principle, then thought experimenters are right out of business. I think both of these claims are wrong, especially the later.

TEs use concepts at hand. Often these are folk concepts, but they needn't be. TEs are frequently constructed at a high level in physics with very sophisticated concepts. They would not be intelligible to the untrained (the folk) and might only be understood after years of study. Rather than thinking that folk concepts are required, it would be better to say that *familiar concepts* are required and that this can include highly sophisticated concepts that have been internalized by the thought experimenter so as to be second nature.

The second point, which is Dennett's, not Nenad's, is that folk concepts are misleading or useless. Not so. Some folk concepts might turn out to be of great scientific value. For instance, Galileo's ship example uses everyday concepts about the motion of a ship and our typical experience inside and outside the ship. It lead to the principle of relativity in both Newtonian physics and special relativity. For a second example, consider Turing's analysis of computability. His account of what is now known as Turing machines is often said to be a thought experiment. I won't describe it here except to say that a very simple, readily understood mechanism leads to some spectacular results. One of these is that most functions are not computable. In both of these cases, folk concepts have given us spectacular results that we now consider among our most sophisticated beliefs.

Nenad's new book is rich in detail and powerful in defence of mental models. His mental models account has become one of the most important and influential accounts of TEs, arguably more popular than any other. *Thought Experiments* will reinforce this opinion. It is a richly rewarding contribution to our better understanding of TEs in particular and how we learn about things in general.

Unfortunately, I must end on a sad note. Recently Nenad died. He was a wonderful friend, interested in everything and with an opinion about it no matter what it was. Every discussion was lively, funny, and included a touch of scurrilous gossip. We shared a seriously left-wing outlook and shared similar views on religion and current politics. Most of all I shall miss future discussions on thought experiments. As I already noted, he (and Nancy Nercessian) were the first to propose the very popular and plausible mental models view of thought experiment. He was particularly insightful on political thought experiments. Again and again I found myself persuaded and always looked forward to the next encounter. The loss is hard to fathom.

References

Miščević, N. 1992. "Mental Models and Thought Experiments." *International Studies in the Philosophy of Science* 6: 215–226.

Miščević, N. 2022. *Thought Experiments*. Cham: Springer.

Nercessian, N. 1993. "In the Theoretician's Laboratory." In D. Hull, M. Forbes, and K. Okruhlik (eds.). *PSA 1992*. East Lansing: Philosophy of Science Association, 291–301.