Croatian Journal of Philosophy
Vol. XXIV, No. 71, 2024
https://doi.org/10.52685/cjp.24.71.3
Received: September 19, 2022
Accepted: June 4, 2024

Miščević On Thought Experiments

DAVID DAVIES
McGill University, Toronto, Canada

I address two claims that Miščević makes in his book Thought Experiments. The first claim is that literary fictions belong to the broader category of what he terms "Imaginative Enactments in Thought" (IET's), but are not TE's properly understood. The second claim is that TE's are indispensable to analytic philosophy. Both claims appeal to Miščević's discussion in the opening chapter of what it is for something to be a TE. I argue for the following conclusions: (1) If TE's are defined in the way that Miščević proposes, then there can in fact be (and indeed are!) works of literary fiction that qualify as TE's. (2) If TE's are defined in this way and are explained in terms of mental models, then whether there can in fact be analytic philosophy without TE's depends upon how we understand the relationship between TE's and counter-factual thinking more broadly construed.

Keywords: Thought experiments; fictional narratives; mental models; analytic philosophy.

Foreword

It is very sad that Nenad's untimely passing has deprived us of what would, I am sure, have been his very lively responses to these papers exploring themes in his wonderful book *Thought Experiments*. But I am very pleased to include, in this commemorative issue of the *Croatian Journal of Philosophy,* a brief paper that celebrates some of Nenad's insightful and valuable contributions to the literature on thought-experiments, contributions that I, like many others, have learned from and drawn upon in my own work.

1

The centrepiece of Nenad Miščević's very interesting book *Thought Experiments* is the further elaboration and defence of his 1992 account of how we are able to learn from thought experiments (TE's) in both science and philosophy. Carving out a middle ground between the pessimistic views of the empiricists—where the best we can hope to get from thought experiments is deductive arguments in sheep's clothing—and the heady views of the Platonists, Miščević has argued that, when we "run" a TE, we are able to activate and draw upon unarticulated and/or unarticulable cognitive resources, some of these innate and some the unarticulated residue of our experiential engagements with the world. He draws here upon Johnson-Laird's idea (1983) that the construction of "mental models" is a crucial part of our comprehension of narratives.

But Miščević's book advances his earlier thinking on these matters in at least two important ways. First, stressing the analogies between real experiments and TE's, he analyses the cognitive work of a TE into a number of distinct stages. The first five stages incorporate the conception and formulation of the TE, and its initial reception resulting in an "intuition" on the part of the receiver. The further stages incorporate processes of (a) "intuitive induction", where we gauge the more general import of the TE through comparison with other related TE's, and (b) seeking "reflective equilibrium", where the import of the TE is determined by locating it in the broader framework of our understandings of the world. Citing Stevin's famous "chain" TE, Miščević notes that "scientists, philosophers and teachers know that [engaging with the narrative] is not the end of the story: one can and should vary the story and generalize the result, and then test the intuition and generalization, comparing them to other spontaneous intuitions and generalizations, or even to information from psychology of belief-formation" (Miščević 2022: 9).

This analysis in terms of stages serves two roles in Miščević's response, in chapter 6, to the challenges to the cognitive status of TE's that have come from experimental philosophy. First, although Miščević does not stress this point, it seems to follow that the intuitions evoked by TE's have cognitive value only when the TE's are elements in the kind of broader investigative practice that the "stages" model describes. Second, analysing the workings of TE's in terms of the "stages" model allows us to identify different places where our intuitions might be untrustworthy and, thereby, to consider measures that might render TE's more epistemically reliable. Both of these points are of special importance for analytic philosophy, Miščević maintains, because TE's are indispensable for the latter. Finally, Miščević argues (chapter 5) that, to properly understand how TE's work in philosophy we need to view them diachronically, as the means whereby philosophical thinking in a given field may develop through engagement with and development of a powerful TE. He develops this point at some length, taking as his principle example the manner in which thinking in the philosophy of language and of mind has developed in different ways in response to Putnam's original Twin Earth TE's.

2

I have always found the "mental model" view of TE's in its various incarnations an attractive one. It preserves, with philosophically modest resources, our sense that TE's can have genuine cognitive value. It also solves nicely Kuhn's puzzle (1964) as to how we can acquire new knowledge of the world without new empirical input. We can do so, it is claimed, because, in constructing a mental model in our comprehension of the narrative of a TE, we are able to draw on otherwise inaccessible understandings of the world that we already possess. I think Miščević does an excellent job of deepening and expanding his earlier published defence of the "mental model" account in this book. So I shall not be questioning Miščević's general positive account of TE's.

What I do want to address, however, are two further claims that Miščević makes, one in the opening chapter of the book and the other in his account of the role to be accorded to TE's in philosophy. The first claim is that literary fictions belong to the broader category of what he terms "Imaginative Enactments in Thought" (IET's), but are not TE's properly understood. The second claim is that TE's are indispensable to analytic philosophy. Both claims appeal, directly in the first case and indirectly in the second, to Miščević's discussion, in the opening chapter, of what it is for something to be a TE. My two critical reflections will take this discussion as premise and argue for the following conclusions:

(1) If TE's are defined in the way that Miščević proposes, then there can in fact be (and indeed are!) works of literary fiction that qualify as TE's.

(2) If TE's are defined in this way and are explained in terms of mental models, then whether there can in fact be (analytic) philosophy without TE's depends upon how we understand the relationship between TE's and counter-factual thinking more broadly construed, an aspect of Miščević's account of TE's that perhaps needs further clarification.

3

In specifying what he takes to be the constitutive features of a TE, Miščević contrasts his own view with Mach's somewhat expansive account. According to Mach, "the planner, the builder of castles in the air, the novelist, the author of social and technological utopias is experimenting with thoughts; so, too, is the hard-headed merchant, the serious inventor and the enquirer" (Mach 1976: 29; cited in Miščević 2022: 10). Miščević does not question the interest of this grouping, but proposes that we view it as a broader genus—"Imaginative Enactments in Thought"—of which "strict TE's" of the sort that we find in science and philosophy are a species. The latter "have as their primary purpose increase of knowledge" whereas the other kinds of IET's listed by Mach have "a different primary motivation".

One kind of IET that Miščević wishes to exclude from the class of strict TE's is works of narrative fiction such as novels and films. He cites my piece on "Art and Thought Experiments" in the *Routledge Companion to Thought Experiments* as following Mach in using the term TE "in a very wide sense" so as to include such artistic fictions (Miščević 2022: 11). He then argues that, while the latter may have *some* cognitive function, their primary function will be either to achieve artistic ends of an expressive or formal nature or to induce enjoyment or other kinds of affect. In defence of his exclusion of artistic fictions from the realm of strict TE's, he further claims that, in such fictions, "the requirements of strictness are weaker than in TE's. In science and philosophy the TE should have a clear and univocal goal, and the proposal that is tested by it has to be decided in a non-ambiguous way. In a literary work ambiguity is often praised as a goal" (Miščević 2022: 11).

Let me note first that, in my piece in the *Companion*, far from following Mach's profligate employment of the term "thought experiment", my use of the term agrees in all essential respects with Miščević's. My aim in that piece was to assess the extent to which—as other authors such as Catherine Elgin (2007), Noel Carroll (2002), and James Young (2001) have claimed—at least some artistic fictions *meet* Miščević's requirements for being strict TE's. According to these authors, at least some literary or cinematic fictions are IET's whose primary intended purpose is to increase our knowledge or understanding. The authors in question further claim that, as a result, at least some works of artistic fiction have significant cognitive value. They thereby espouse some form of what is usually termed "literary cognitivism". In my piece, drawing on a couple of earlier articles (Davies 2007, Davies 2010), I argue that the first claim is correct but express significant reservations about the second claim.

These reservations obtain because a defender of literary cognitivism must meet certain empiricist challenges analogous to those that Miščević surveys in his overview of empiricist criticisms of the cognitive credentials of TE's in science. A representative sample of the kinds of challenges confronting the literary cognitivist can be found in Jerome Stolnitz's paper (1992) "On the cognitive triviality of art". Stolnitz begins by suggesting that we cannot learn anything interesting about the world through reading works of fiction because the supposed "truths" in such works are generally banal and imprecise. All we might hope to learn from reading Jane Austen's *Pride and Prejudice*, for example, is that "stubborn pride and ignorant prejudice keep attractive people apart," and even here it is unclear what the scope of this claim is. To the response that this fails to do justice to the general truths about the world that may be gleaned from works of fiction, Stolnitz responds that, even if there *were* genuinely interesting truths about the world exemplified in the narratives of works of literary fiction, we couldn't *learn* those truths in our engagements with those works of fiction because the work provides no *empirical support* for such putative truths. All we are given in the fictional narrative is a single non-real example which has been gerrymandered to make those "truths" apparent. Echoing empiricist critics of TE's in science, Stolnitz maintains that the best we might get from reading literary fictions is interesting hypotheses that might then be subjected to independent empirical test.

The literary cognitivists cited above respond to this kind of challenge by arguing that at least some literary works function as extended TE's, and can therefore share in the kinds of cognitive value ascribable to TE's in science (Elgin) and philosophy (Carroll). In critically discussing this strategy on the part of literary cognitivists, I have pointed out (see especially Davies 2010) that the strategy can serve cognitivist aims only if we counter the empiricist criticisms of TE's in the latter domains. In fact, a model of TE's like the one defended by Miščević seems to be just what the literary cognitivist needs. If the running of a scientific or philosophical TE can yield genuine knowledge of the world—without the need for independent empirical testing—because the TE draws on genuine but unarticulated, or unarticulable, cognitive resources, then, if literary fictions are TE's, surely the same can apply to them, and Stolnitz's objections are answered.

Literary cognitivists have generally assumed that their case is made once it is granted that some literary fictions are TE's, but even if one supplements the cognitivist's case with something like a "mental model" account of TE's, there are still issues that need to be addressed (see again Davies 2010). Miščević's "stages" model, in fact, provides further reason to be sceptical about the literary cognitivist's claims, since the consumption of literary fictions does not seem to be part of a larger practice of consuming and testing TE's, and it is, according to this model, the location of our running of TE's within such a practice that confers cognitive credibility upon the intuitions they evoke.

But the issue of present concern is whether at least some works of literary fiction can meet Miščević's requirements to count as "strict TE's", and here I think the answer must be a positive one. The requirement, we may recall, is that the principal aim of the narrative be a cognitive one: the primary purpose should be to increase knowledge, and, with this in mind, the "lesson" of the TE should be clear and not trade in ambiguity. Perhaps fittingly, we can show that this requirement can be met by means of a (philosophical) TE! Let us imagine two literary authors—let us call them Edward and Graham. Suppose that Edward, in a literary essay, expresses the view that our moral duties to our friends should outweigh our moral duties to our country. When

Graham hears of this, he strongly disagrees and undertakes to demonstrate how duty to country can, at least on occasion, outweigh duty to friends. He does so by writing a literary fiction where, when the reader grasps the genuinely conflicting nature of the duties to friend and country confronting the main protagonist, her intuitions will accord with those of the protagonist when the latter decides to weight duty to country over duty to friend. The motivation for composing the fictional narrative in this case is clearly cognitive, and there is no attempt to make the situation ambiguous in any relevant respects. Thus, by Miščević's criteria, we have a work of literary fiction that is a strict TE.

In fact, we do not need to appeal to a TE to make this case. For, at least on some accounts, what we have described in hypothetical terms was what actually led Graham Greene to write his novel *The Third Man* (1950) to counter a claim about how to balance moral duties to friend and country voiced by E. M. Forster in his essay "What I believe" (Forster 1938/1951). And it is not difficult to find other examples of works of literary fiction whose primary aim is cognitive in this way. The original edition of Anthony Burgess's *A Clockwork Orange* (1962), for example, is an extended IET intended to explore the moral issues surrounding the treatment of social deviance. Here again the purposes motivating the construction of the narrative are clearly cognitive in the manner required by Miščević. But it is *not* sufficient to meet Miščević's criteria that the author of a literary fiction works with the elements that define a philosophical issue: Carl Reiner's film *All of Me* arguably takes as its basis the kinds of hypothetical cases that drive debates about the place of embodiment in our sense of personal identity, but the aim of the film is clearly to entertain rather than enlighten the viewer (for a discussion of this case, see Smith 2006).

We thus have examples of existing literary fictions that (1) have as their primary purpose the increase of knowledge or understanding, (2) are not intentionally ambiguous, and (3) are, if Johnson-Laird's "mental model" account of narrative comprehension is correct, comprehended through constructing a mental model. They thereby fit Miščević's description of a "strict" TE in chapter 2 of his book: "We have characterized a TE as a process that starts with a design, which involves the determination of the goal(s), in particular the thesis/theory to be tested, and the construction of a scenario to be considered. We noted that it then proceeds with the presentation of the scenario thus constructed to the experimental subjects. On the side of the subject, the experiment then continues with the typically imaginative contemplation of the scenario plus some piece of reasoning, culminating in the decision ("intuition") concerning the thesis/theory to be tested."

4

In the final section of this paper, I want to at least raise some questions about Miščević's claim that TE's are "indispensable" for analytic philosophy. We find an argument for this claim, at least with respect to practical philosophy, in the following passage: "The traditional sources of insight here are either facts (including presumed facts), principles or TE's. Facts are useful and indispensable, but taken alone they don't teach us about what is valuable, morally prohibited, morally indifferent and so on. We need at least principles. But how do we test principles? The only source here are intuitions and the indispensable testing grounds are TE's" (Miščević 2022: 26). As he later puts this, for philosophy "TEs are indispensable. Philosophy does not use [a] laboratory to test its theories; the only experiments available here are those in thought....Although life without TEs might be possible for science, it is practically impossible for philosophy" (Miščević 2022: 87, 98).

We might reformulate this argument as follows:

(1) The claims that philosophers seek to evaluate are *modal* in the sense that they are not just claims about how things actually are but about how things must be, or can't be, or ought to be.

(2) To evaluate a modal claim, we need to engage in counter-factual reasoning.

(3) To engage in such counterfactual reasoning is to entertain a thought experiment.

(4) So philosophy cannot do without TE's.

Points (1) and (2) seem valid if we restrict ourselves to attempts to *defend or establish* a modal claim. To defend a general modal claim is to maintain that it would lead to the right results in possible as well as actual cases, and to assess a possible case requires counterfactual reasoning. It might seem that the cases brought *against* such a claim could be actual cases and would therefore not call for counterfactual reasoning: in countering the claim "all A's must be B", we might point to an actual A that is not B. It might be responded that we will still need counterfactual reasoning to establish that we have a genuine counter-example to the universal claim. But rather than pursue this issue, I want to look at the move from (2) to (3) and (4).

As we saw, Mach understood the idea of a TE very broadly—it includes any process of working out in one's head how to proceed in a given instance, where this necessarily involves considering various options and thus counter-factual reasoning. On this account, the merchant in the market who deals with a customer trying to haggle for a cheaper price is engaged in a TE. Miščević is critical of this broad construal of TE's, but this is on the grounds that a TE must have a primarily cognitive purpose. But does Miščević hold that, *as long as this further condition is satisfied*, any instance of counter-factual reasoning is a TE? Suppose we term such a view the "cognitively motivated counterfactual reasoning" (CMCR) view of TE's. While the CMCR view seems required if (4) is to follow from (1) and (2), it also raises a number of questions:

(i) The CMCR view will incorporate many examples of counterfactual reasoning in philosophical and other contexts that we would not normally think of as (philosophical) TE's of the sort discussed throughout Miščević's book. This *broad* conception of TE's would resemble the one that Miščević ascribes (2022: 43) to Buzzoni according to which "TEs are the condition of the possibility of REs because, without the a priori capacity of the mind to reason counterfactually, we could not devise any hypothesis and would be unable to plan the corresponding RE that should test it" But Miščević seems sceptical about Buzzoni's approach.

(ii) If Miščević is operating with the CMCR view of TE's, it is difficult to make sense of his sympathetic response to Williamson's account, which clearly rejects the CMCR view. Indeed, both Williamson (2016) and Miščević seem concerned to distinguish TE's from cognitively motivated counterfactual reasoning more generally. Miščević cites here Williamson's discussion of the hunter who deliberates about whether to attempt to ford a stream by jumping across it at its narrowest point. What distinguishes such a case from counter-factual reasoning more generally, for Williamson, is the hunter's use of imagination, something that cannot be replaced by more abstract reasoning. Miščević develops this idea by proposing that the imagination here serves a particular role, namely, the construction of a mental model of the counterfactual situation. On the mental-modelling approach, TEs are sophisticated "re-modellings in the head" whose most important feature "is precisely their concrete and quasi-spatial character" (Miščević 2022: 47). This strongly suggests that for Miščević only those cases of counter-factual reasoning that have these distinctive features of mental modelling count as TE's, contrary to the CMCR view. But in this case, it seems, we cannot derive (4) from (2).

(iii) However, certain other remarks by Miščević seem to place him closer to the CMCR view. For example, in discussing the distinctive features of mental models, he states that "TE's might involve language-like representations and inference and computation on them, but *typically*, they involve more concrete representations, such as are used in imaginative operations" (Miščević 2022: 53, stress added). This seems to erase the distinction that Williamson is trying to draw in his appeal to the use of the imagination in TE's as contrasted with other more formal kinds of counter-factual reasoning. On the other hand, in another puzzling remark which seems to indicate a departure from the CMCR view, Miščević claims that a TE need not involve counter-factual reasoning because some cases considered in a TE can be real" (Miščević 2022: 46). One wonders here whether, for such cases to count as TE's, they must in fact involve counter-factual reasoning about the real case. If not, why think of them as TE's rather than imaginative engagements with an actual case, as occurs in the mental modelling of a non-fictional narrative. Also, this seems to conflict with the claim that "thought-experimenting involves proposing and considering counterfactual scenarios (Miščević 2022: 44).

These are issues upon which I am sure Miščević would have provided further clarification and enlightenment had he been able. But they are issues that only present themselves because of the intellectually engaging aspects, as described earlier, of Miščević's overall enterprise in this very interesting book.

References

Burgess, A. 1962. *A Clockwork Orange*. London: William Heinemann.

Carroll, N. 2002. "The Wheel of Virtue: Art, Literature, and Moral Knowledge." *Journal of Aesthetics and Art Criticism* 60 (1): 3–26.

Davies, D. 2007. "Thought Experiments and Fictional Narratives." *Croatian Journal of Philosophy* 19: 29–46.

Davies, D. 2010. "Learning through Fictional Narratives in Art and Science." In R. Frigg and M. Hunter (eds.). *Beyond Mimesis and Convention: Representation in Art and Science*. Boston Studies in the Philosophy of Science 262. Dordrecht: Springer, 51–70.

Davies, D. 2018. "Art and Thought-experiments." In M. T. Stuart, Y. Fehige, and J. R. Brown (eds.). *The Routledge Companion to Thought Experiments*. London: Routledge, 512–25.

Elgin, C. Z. 2007. "The Laboratory of the Mind." In W. Huerner, J. Gibson, and L. Pocci (eds.). *A Sense of the World: Essays on Fiction, Narrative, and Knowledge*. London: Routledge, 43–54.

Forster, E. M. 1951 [1938]. "What I believe." In *Two Cheers for Democracy*. New York: Harcourt Brace.

Greene, G. 1950. *The Third Man, and the Fallen Idol*. London: William Heinemann.

Johnson-Laird, P. N. 1983. *Mental Models*. Cambridge: Harvard University Press.

Kuhn, T. 1964. "A Function for Thought Experiments." Reprinted in *The Essential Tension*. Chicago: University of Chicago Press, 1977, 240–265.

Mach, E. 1976. *Knowledge and Error*. Dordrecht: Reidel.

Miščević, N. 1992. "Mental Models and Thought Experiments." *International Studies in the Philosophy of Science* 6 (3): 215–226.

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

Smith, M. 2006. "Film Art, Argument, and Ambiguity." *Journal of Aesthetics and Art Criticism* 64: 33–42.

Stolnitz, J. 1992. "On the Cognitive Triviality of Art." *British Journal of Aesthetics* 32 (3): 191–200.

Williamson, T. 2016. "Knowing by Imagining." In G. Currie (ed.). *Knowing Through Imagination*. Oxford: Oxford University Press, 113–126.

Young, J. 2001. *Art and Knowledge*. London: Routledge.