VERSITA

HUMAN AFFAIRS *22*, 448–453, 2012
DOI: 10.2478/s13374-012-0036-x

EPISTEMIC PING-PONG: A CRITICAL REVIEW OF MAREK PICHA'S THOUGHT EXPERIMENTS[1]

RADIM BĚLOHRAD

Marek Picha's recently published book, *Kdyby chyby: Epistemologie myšlenkových experimentů* [*Woulda, coulda, shoulda: epistemology of thought experiments*] is a remarkable attempt at investigating the problems of thought experiments—an area that has not been systematically dealt with in the Czech philosophical context. Thought experiments are epistemic tools that are frequently used in science and, in particular, philosophy, but the relevance and strength of the beliefs that are arrived at in the process of thought experimenting is seldom considered more closely. One camp of scholars embraces thought experiments almost as naturally as they do deductive arguments or empirical observation, others reject them outright as mere fantasizing. Assuming that a substantial proportion of philosophy is based on thought experiments, the question of the limits of their legitimacy is an acute one.

Besides a foreword, an introduction and a conclusion, Picha's book consists of five dense chapters supplemented by a list of selected well-known and representative thought experiments.

The opening chapter attempts to define thought experiments. Picha distinguishes between a *wide concept*, according to which thought experiments are forms of reasoning that include a hypothetical scenario, its deliberation, the resulting belief and further argumentation based on it, and a *narrow concept,* which only consists of the deliberation of a particular hypothetical scenario, the other aspects being only the *forms of use* of the thought experiment. The weakness of the wide concept, Picha believes, is that one can arrive at a great number of beliefs using the same hypothetical scenario, but the wide approach is bound to treat these processes as different experiments. To clarify the identity conditions, Picha introduces the concept of *distillate*—a kind of reconstruction excluding the unimportant details and

[1] Picha, M. (2011). *Kdyby chyby. Epistemologie myšlenkových experimentů.* Olomouc: Nakladatelství Olomouc.

retaining the structurally-conceptual core. This enables him to define the identity of thought experiments by means of the identity of the distillates. What a distillate captures is a set of instructions for determining which particular situation to imagine in order to find something out, which, in turn, is Picha's definition of a thought experiment.

The second chapter focuses on the role of thought experiments. Not needing to defend the obvious didactic function, Picha quickly moves to the controversial epistemic function. This is the question of whether thought experiments are capable of delivering new information, information that the experimenter does not already possess *in some form*. Picha spends some time reviewing the various meanings of *new information* and concludes that what is really at stake is whether the beliefs that thought experiments produce are *informative* and perhaps *surprising*. Even though the answer to this question is positive, this is not the key question. What the reader is keen to know is whether the beliefs produced by thought experiments are *justified*. Picha presents five optimistic positions: Brown's apriorism, Mach's psychologism, Kuhn's conceptualism, Nersessian's mental modeling theory, and his own associationism. Picha rejects apriorism on metaphysical and epistemological grounds, and conceptualism due to its epistemic irrelevance, whereas psychologism and mental modeling are treated as respectable rivals. However, Picha believes that associationism is preferable to the two rivals, as it can make better sense of the experimenter coming to contradicting beliefs in the repeated application of the same thought experiment. I shall return to this argument below.

Somewhat surprisingly, after defending associationism in the second chapter, Picha goes on to claim that his favorite concept of thought experiments is as useless as the rival concepts in deciding the question of epistemic justification. He believes that epistemic justification can be settled regardless of the particular theory of thought experiments. That is the goal of the third chapter.

The third chapter presents an argument for the conclusion that thought experiments can only provide a limited degree of justification for beliefs—*prima facie* justification (pf-justification). They cannot be considered trustworthy sources of knowledge, however. First, Picha polemicizes with Thomas Senor about the definition of pf-justification and offers his definition: a belief is pf-justified if there are epistemically relevant reasons for the belief and the belief has not been refuted. Epistemic relevance and refutation are explained in subsequent sections. Next, Picha sides with the skeptic and rejects the argument for the reliability of thought experiments. None of the theories of thought experiments presented in chapter two (including Picha's favorite associationism) can meet the generality problem for reliable epistemic processes. However, Picha believes that a limited epistemic role for thought experiments can be defended: even though they are not reliable, thought experiments are still epistemically relevant, because they can lead to the acceptance of beliefs, and if the beliefs have not been refuted, the thought experiments constitute pf-justification for the beliefs. This gives thought experiments their dialectical function: if a belief is pf-justified by means of a thought experiment, the burden of proof shifts to the opponent, whose turn it is to assess the belief. A brief but important part of the third chapter is Picha's treatment of another philosophical tool—intuitions, defined as underived beliefs. These share the epistemic status of thought experiments, being a source of pf-justification.

In the fourth chapter Picha poses the question of whether the role of thought experiments is irreducible. He outlines the challenge of *eliminativism* (as presented by Norton), according

to which all thought experiments can be reconstructed as arguments based on explicit or hidden assumptions, and provide no more justification for the resulting beliefs than the arguments do for their respective conclusions. Then, making use of Galileo's Pisan experiment, Picha demonstrates and analyzes two objections to eliminativism—Brown's and Gendler's—and finds them both inadequate. The dispute turns on the correct argumentative reconstruction of the Pisan experiment. Picha distinguishes between *exemplary* and *illustrative* thought experiments and contrasts the Pisan experiment with Parfit's *Division.* Then, he claims that Galileo's experiment can really be replaced by an argument as it is an instance of the illustrative type of thought experiment, whose role is to strengthen an already accepted general belief. In contrast, Parfit's *Division* is more plausibly interpreted as an exemplary thought experiment whose goal is to persuade us to accept a new general belief. Even though the difference between illustrative and exemplary thought experiments is contextual, Picha makes his case for the existence and irreducibility of exemplary thought experiments.

The final chapter is the most practical of all as it deals with the criteria of assessment of thought experiments. Picha carefully distinguishes between errors of thought experiments and errors of arguments utilizing thought experiments, which mistakenly discredit thought experiments themselves. In the former category one can find the errors of *epistemically impossible conclusion, unreliable conclusion,* and *unreachable conclusion.* The latter contains the frequent errors of *irrelevance, incorrect generalization,* and *impatience.* In the course of the chapter Picha touches upon the challenges of experimental philosophy and the study of introspection, and argues that they do no more than warn us to be cautious when dealing with thought experiments and intuitions. The chapter concludes with a set of *critical questions* that help assess arguments by thought experiment.

Let me now turn to some critical points. My first reservation concerns the way Picha treats the rival optimistic theories of thought experiments in chapter two.

First and less importantly, as the different theories are presented, one feels as if they were mutually exclusive conceptions of (among other things) the activity that we engage in when we carry out thought experiments. However, there is no reason to suppose that there could not be various ways of proceeding when drawing conclusions from the contemplation of hypothetical scenarios. In some cases the underlying process may be one of actualizing an implicit or unconscious belief, in others it could be the process of remembering and analogical thinking. Picha claims that associationism can be empirically falsified. But if there are various patterns of thought experimenting, the fact that the regions responsible for episodic or semantic memory do not light up in fMRI does not mean that thought experimenting is not going on.

More importantly, I do not believe that associationism is really preferable to its best rivals—psychologism and mental modeling theory. Picha claims that his theory can explain more easily how "one experimenter can come to contradicting conclusions in the repeated application of the same thought experiment" (p. 63). This presupposes that we are clear about the *sameness* of thought experiments, which takes us back to chapter one, where this notion is defined by means of the sameness of distillates. The problem is that the notion of the sameness of distillates is no clearer. There might be a very fine-grained criterion grounding a distillate in its structure plus the particular *expressions* filling the structure, or a

coarse-grained criterion identifying the distillate with the structure plus the *concepts*. Perhaps not even the structure is essential. The problem is that we simply do not know and have to rely on an intuitive grasp of the sameness of thought experiments, just as Picha does when he uses Williams's *Mad Surgeon* experiment in its two formulations and claims that they "intuitively…seem identical" (p. 64). There are, however, good reasons to suppose that we are dealing with two different experiments. Williams's experiment, very roughly, describes a case of mind-body swap in two formulations. One formulation results in the belief that I go where the mind goes and the other suggests that I stay with the body. The problem is that Williams's use of personal pronouns in the two descriptions of the hypothetical scenario makes us associate ourselves with the mind or the body *in advance* and, thus, the descriptions beg the question. What is important for our purposes, however, is that even if we use the coarse-grained criterion, the concept of *I-the body*, implicit in one of the descriptions, is certainly different from the concept of *I-the mind,* presupposed by the other, and the two formulations are, therefore, different thought experiments. And if they are different, the explanation as to how one can come to contradicting conclusions is, I suppose, no deep mystery.

My second qualm has to do with Picha's polemics regarding Senor's concept of *prima facie* justification. For Senor, a belief is pf-justified if it is the result of an epistemic process capable of delivering *ultima facie* justification (i.e., an epistemically reliable process), even though there may be a stronger reason overriding the belief. So if Alice thinks she sees Bert in the distance and has dispositional knowledge that Bert is away in Paris, her belief that she sees Bert is pf-justified despite her other belief, because it is the result of a reliable epistemic process—perception. What Picha finds hard to accept is the idea that an overridden belief can still be justified and he proposes a definition according to which a belief is pf-justified if and only if there are epistemically relevant reasons for it and *it has not been overridden*. In other words once you arrive at an overriding belief, you cannot call your former belief pf-justified.

I believe Picha's case against Senor is a "straw man." In my opinion Senor is not bound to accept the existence of overridden pf-justified beliefs. According to Senor's definition, a belief is pf-justified if it is brought about by a reliable process, and it is irrelevant whether it is later overridden or not. One reason why Senor need not include the overriding clause in the definition is that, normally, the moment a belief is overridden by a stronger contradictory belief, it ceases to exist. There is no danger that it would somehow continue to exist and the attribute *pf- justified* would still be applicable to it. This shows why Picha's extra condition that a pf-justified belief cannot have been overridden is redundant.

Let me now turn to another of Picha's objections to Senor's account. Allegedly, Senor has to allow for the possibility of a person having two contradictory beliefs both of which are pf-justified. Consider Alice again. She has an actual perceptual belief that she sees Bert and a dispositional knowledge that Bert is in Paris. Her perceptual belief is pf-justified, since perception is a reliable epistemic process. Her knowledge is, of course, uf-justified, and, therefore, according to Senor's definition pf-justified as well. Both of Alice's beliefs are thus pf-justified but contradictory. And that, according to Picha, is a problem.

In the assessment of this case it is the *actual* and *dispositional* attributes that are important. A dispositional belief is one that a person has, but is not currently aware of. It is a belief that one can, on reflection, draw from the stock of accepted beliefs. Each person has

a great number of mutually inconsistent dispositional beliefs, perhaps because these beliefs come from contexts which the person has never checked for mutual compatibility. Picha believes that having such beliefs is quite natural. So why would having two inconsistent pf-justified dispositional beliefs be a problem? The simple fact that they are dispositional, i.e. they are not currently the focus of my conscious awareness, explains how they can be mutually inconsistent. If, moreover, they were both caused by processes capable of uf-justification, but perhaps on very different occasions, it is only natural that the person retains both of the beliefs until one of them is explicitly confronted with an overriding reason and ceases to exist.

The same holds for the combination of an *actual* pf-justified belief and a *dispositional* pf-justified belief (Alice's case). Again, since one of the beliefs is dispositional, the person may not be aware of any contradiction. Once the contradiction is seen, one of the beliefs is bound to cease to exist. The only situation that would present a problem for Senor is having one actual pf-justified belief and one contradictory actual uf-justified belief. If Alice actually thinks she sees Bert and at the same time is aware of the fact that she knows Bert is in Paris, then there is something wrong with her. But there is no reason to suppose that Senor's definition leads to such an unacceptable distribution. As soon as Alice realizes fully that Gert is in Paris, that is, as soon as her dispositional knowledge turns into actual knowledge and starts competing with her pf-justified perceptual belief, the latter does not turn into a pf-justified overridden belief, as Picha claims, but ceases to exist. So the problematic co-occurrence of two prima (ultima) facie justified actual beliefs does not ever need to occur (putting aside the irrelevant case of mental illness).

To be fair to Picha, he realizes the mistake several pages later (p. 73) when he says that according to both his and Senor's conception, an overridden pf-justified belief is not pf-justified since it is not a belief any longer. But then the preceding six pages of the chapter are pointless and the only argument that Picha can put forward against Senor's is the one to which I turn now.

Picha claims that Senor's conception has unacceptable consequences when we consider a third-person assessment of someone's belief. Suppose Alice has the pf-justified belief that she sees Bert, and her colleague Cecil knows that Bert is in France. Picha asks whether Alice's belief is justified from Cecil's perspective. According to Picha's reading of Senor, he is bound to consider that the belief is justified even from Cecil's perspective, because it has been caused by a reliable process. But Picha believes that the belief is unjustified from Cecil's perspective, as he has overriding counter-evidence. The situation is tricky, because the contradicting beliefs are now distributed between two people. One, Alice, has the positive perceptual belief, but not the negative overriding belief. The other, Cecil, knows the overriding evidence, but does not possess the perceptual belief.

Still, I do not see what exactly is wrong with the contention that Alice's belief is justified from Cecil's perspective. First of all, what does it mean to say that a person is justified in holding a belief? To me, it means that the person followed the best available evidence (including a reliable epistemic source) when adopting the belief. What it does not mean, however, is that the belief is actually true. Justification does not entail truth. I can be justified in believing it is 6 o'clock by looking at the clock above my door, which in most cases shows the right time, even though it may actually be 4:30. Moreover, if justification entailed truth,

the predicate *true* in the definition of knowledge as true, justified belief would be redundant. But if that is so, there is no problem in Cecil saying: "I know her belief is false, but she is justified in holding it." I think the problem lies in a misunderstanding about what the concept of justification applies to. The formulations Picha uses suggest it is beliefs that are justified or unjustified. But, in my opinion, this use is only derivative. In fact, it is primarily people who are justified in holding beliefs. If Jack believes p for good reasons, Jill believes non-p for good reasons, and Joe withholds believing in p for good reasons, is p justified, unjustified or something else? But if Jack is justified in holding a belief, Jill might be absolutely certain the belief is false and still say that for all Jack knows, he is justified in holding the belief. In sum, I do not wish to put forth the stronger claim that justification is not an absolute notion. All I want to point out is that there is a sense of justification that ties the concept to the person holding a belief, and if that is so, whatever other people know about the matter, that person might still be justified in holding that belief.

At the beginning of the book Marek Picha admits that the theme of his book is not particularly exciting. He deserves all the more credit for doing such a good job mapping the various theories, being able to assess them and defending a really subtle position. I only wish the conclusion of his book could be more optimistic. If thought experiments and intuitions provide at most *prima facie* justification, and if *prima facie* justification can at most shift the burden of proof, some considerable areas of philosophy that rely heavily on thought experiments and intuitions (such as personal identity theory) come to no more than a kind of epistemic ping-pong with little hope of reaching knowledge.

Department of Philosophy,
Faculty of Arts,
Masaryk University,
Arna Nováka 1/1,
602 00 Brno
Czech Republic
E-mail: belohrad@phil.muni.cz