

## **Review of dissertation**

**Title: "Epistemological Externalism in Mental Models"**

**Author: Plamen Chergarov**

The dissertation consists of approximately 216 standard pages, divided into an introduction, five chapters, a conclusion, and a bibliography with an impressive number of sources, primarily in English.

The author's goal is to defend the position of epistemological externalism against internalism using contributions from the philosophy of mind, psychology, and neuroscience. In other words, the text aims to solve an epistemological problem using contributions drawn from other disciplines.

Chergarov's argument relies on the assumption that physicalism within the metaphysics of the mind is the correct position— "the identity theory is the best explanation for the nature of consciousness" (p. 39). This position allows the author to introduce the concept of the "mental model," which is central to the dissertation's argument. The mental model is defined as internalized by a subject (p. 72), as a "representation" isomorphic to the neural structure (p. 84) that "allows interaction with the world by organizing incoming information in a meaningful way," where "the taken information is part of the mental model even though it is external to it" (p. 77). Since human knowledge as, for instance, a justified true belief, according to the author, is also part of a "mental model" (p. 84), it follows that knowledge is determined by external rather than internal factors (since mental models are essentially "internalized" external factors), thereby redeeming epistemological externalism.

I agree with the author's assertion that externalism is the correct position in the contemporary epistemological debate. I applaud his interdisciplinary approach and the erudition demonstrated in the text. I fully accept the presupposition of physicalism as the correct position within the metaphysics of mind, although I have reservations regarding the identity theory as a specific physicalist position. However, as is appropriate for a dissertation review, I will offer several critiques and objections to encourage Plamen Chergarov's further development.

First, the "cohesion" between epistemology on one hand and psychology and neuroscience on the other is not clearly argued and demonstrated, resulting in a lack of methodological consistency in the dissertation.

In the conclusion, on p. 188, the author points out one of the problems with this "cohesion" - the predicate "is mental" (in "mental model") is retained solely to make his argument relevant to the debate between internalism and externalism. While this is understandable given the naturalistic framework Chergarov adopts, if retaining a predicate for the sake of addressing a philosophical debate is the only justification for its use, does this not imply that the contribution to the debate is somewhat artificial and does not genuinely belong within the premises upheld by the debaters (internalists and externalists)?

Second, it seems Chergarov believes that adopting the framework of naturalism as a common methodological stance and reliabilism as an epistemological position partly compatible with naturalism is sufficient to ensure such "cohesion." According to A. Goldman, reliabilism indeed allows for the justification of beliefs based on the empirical assessment of the reliability of cognitive processes, and it appears Chergarov's argument tries to adhere to this - "mental models" take the place of "cognitive processes" as a specific theoretical construct chosen for a more detailed presentation of the "naturalized component" on which the reliability of beliefs depends and which more clearly shows how this reliability depends on external factors. However, in the last chapter, Chergarov reverses the relationship between epistemology and mental models - mental models are characterized as epistemically valuable because they can be viewed from the perspective of reliabilism in epistemology (as the author summarizes on p. 189). In other words, before this chapter, mental models are used to "naturalize" knowledge in its epistemological treatment, but in this chapter, reliabilism as a position presenting normative criteria for knowledge is used to "normativize" mental models.

Based on these grounds, I believe that although original and interesting, the central argument in the dissertation is methodologically inconsistent, and I cannot unequivocally determine whether it achieves its goal.

Third, the concept of "mental model" does too much work in the text to make it clear what exactly is being summarized through it, even though the author strives to present a clear definition. On p.

84, as previously noted, the author claims that the mental model is "isomorphic to the neural structure." Abstracting from the fact that a theoretical construct from psychology (such as the "mental model") can hardly be isomorphic to the components of nerve cells and/or their spatial and functional organization in the nervous system, this definition implies that the "mental model" is different from the "neural structure," since the relationship between x and y, even if it is a relationship of isomorphism, has the consequence that x and y are not identical. However, in the conclusion, on p. 188, while summarizing his positions, the author claims that "mental models" are identical to "neural models," which essentially contradicts the implication of the statement on p. 84. Therefore, I would like to ask: ultimately, is a mental model identical to a neural model, or is it isomorphic to a neural model (assuming, due to the lack of evidence to the contrary, that "neural structure" is synonymous with "neural model")?

Fourth, although the author considers "mental" as an appendix whose sole role is to make epistemological debates about knowledge "compatible" with contributions from psychology and neuroscience, he seriously considers "mental models" as opposed to the "behaviorist program," i.e., as they are classically viewed in psychology, where their "mentality" is considered an real property. As an example of the historical emergence of the concept of "mental model," Chergarov claims that it appears when "Tolman convincingly demonstrated the inconsistency of the behaviorist program in a series of experiments" (p. 87). This "convincing demonstration" is presented by the author as follows: "[...] rats, freely allowed to run in a maze, perform better when a reward is subsequently added to this maze. This means they have built some internal model of the maze, independent of stimuli, rewards, and learned behaviors" (ibid.). Regarding this statement, I have the following question for the author: given the reconstruction of the experiment where the rats' behavior becomes more efficient after the addition of a reward, to what extent is it logically valid to say that the "internal model" of the rats is independent of stimuli and rewards?

Despite these issues, I highly value Plamen Chergarov's dissertation and consider it one of the quality dissertations developed in the Department of Philosophy. The most attractive feature of the text, namely the author's approach to the problem and the attempt to solve it, stands in clear contrast to the purely reconstructive nature of dissertations defended in the Department. Adding to this the focus on contemporary debates and problems in philosophy and the scientific style of work, I have no doubt that the work meets the requirements for a well-written dissertation.

Therefore, I confidently vote "yes" to awarding the educational and scientific degree "Doctor of Philosophy" to Plamen Chergarov.

May 18, 2024

Assoc. Prof. Dr. Dimitar Elchinov