REVIEW

of the dissertation of Plamen Chergavor on "**Epistemological externalism in mental models**",

submitted for awarding educational and academic degree "doctor" in professional area 2.3. Philosophy,

by Assoc. Prof. Iassen Zahariev Zahariev, New Bulgarian University

The dissertation has a volume of 213 pages. It consists of an introduction, six thematically separated chapters, a conclusion and a bibliography. The bibliography includes 269 titles in Bulgarian and English languages. The dissertation was written under the supervision of Prof. Aneta Karageorgieva, Doctor of Science.

General notes on the structure and content of the dissertation

The author identifies his research as part of the modern paradigm of naturalization of philosophy, including epistemology. It is correctly noted at the very beginning of the abstract, but not in the dissertation, that in this paradigm it is assumed that mental and neural models are identical: "Therefore, mental and neural models are treated here as identical. This identity is assumed in the present work." (p. 1 of the Abstract). It seems to me necessary to point out that accepting as self-evident the truth of this thesis about identity, which is open to discussion in modern philosophy, does not seem to be a good starting point for a philosophical dissertation. The work does not consider the conditions for the validity of this statement and its consequences. Such an approach in the philosophical tradition after Kant is accepted to be called "dogmatic", regardless of what the claims and results of the research are further on.

The first part of the work, marked as Chapter 2 following immediately after the Introduction, provides a detailed summary of the debate on the justification of beliefs. This debate makes sense insofar as the so-called "traditional" definition of knowledge as "justified true belief." The author has devoted dozens of pages referencing the various positions of internalism and externalism. One of the claims of the dissertation is that in the debate about the justification of beliefs, epistemological externalism is more justified than internalism. The deep philosophical point of internalism is that it is justified to claim that one knows something if one has access to the justification of one's own knowledge, that is, one must be aware not only of one's knowledge but also of the grounds of that knowledge. Let's tentatively call it the doctrine of access. The author

shares the view that knowledge can be justified on the basis of so-called "reliable" neuro-biological cognitive processes, even if the knower does not have conscious access to them, which is considered a powerful blow against internalism in favor of externalism. The true belief produced by a reliable process is knowledge, even without needing access to internal justification, though the latter is possible. This part focuses on the differences in the two doctrines regarding the so-called epistemic virtues and on the question of the infinite regress of grounds or justifications of knowledge.

In this spirit, the author comes to the second part of the study, dedicated to mental models that are interpreted externally. The mental model is not necessarily a conscious one, although it is 'inside', probably somewhere in the subject's cranial box, often marked in the dissertation with a capital Latin letter S. It is not quite common to write and speak like this, but it is difficult to understand the prepositions "inside" and "outside" in the paradigm of this epistemology in a different way than this one. Mental models differ from external "cultural" or "mathematical" ones in that they are inside us, somewhere in the folds of the brain. There is no other way to understand this "internalization" (p. 72) except physiologically. In support of his thesis, the author refers to the ideas of Chalmers and Clarke about external consciousness, according to which certain tools - objects and means can be interpreted as part of our own consciousness if they are related in some way to our cognitive activity. According to the main premise of the work, namely that mental and neural models are identical, the author explicitly admits, albeit in parentheses in the abstract, but again not in the text of the dissertation, that he cannot accept the thesis of external consciousness at all. The author assumes that "external elements are part of the increased or decreased competence of our mental models" (p. 15 of the Abstract), a statement that does not have a particularly clear meaning.

Having defined the mental model in general as a representative cognitive structure, a neural map of the world, the author moves on to the influence of areas external to the subject on the formation and functioning of these models. These external fields are respectively culture and technology. It is not entirely clear here why technology is separated as something different from culture. Through numerous examples from various scientific and technological fields, the author shows something inherently banal not only in philosophy, but also in the everyday thinking of people in all eras, namely that the environment and technology as part of this environment have a

defining role in this , which we know. The formulation of this truism through the special terminology of epistemological externalism adds nothing new to it.

The last part of the thesis before the conclusion is about the reliabilistic interpretation of mental models. The latter are seen as values in the context of the so-called "aretic epistemology". According to the author, mental models are reliable mechanisms of cognition, and the models themselves are reliable if they produce a reliable result. Thus, the category of "reliability" in a dissertation is raised to the third degree. Reliable cognition is the result of a reliable cognitive mechanism, which in turn derives its reliability from the reliable output, which is reliable because it leads to knowledge. Attempting to formalize this conundrum on p. 163 in no way makes the concept of "reliability" any clearer or more meaningful. I am tempted to use the author's own words to illustrate this classic circularity of argument: "Knowledge is obtained when the reliability of the mental model is so high that it reflects the world in a way that we would characterize as knowledge." It seems to me that further comment here is unnecessary.

Critical notes and recommendations

The author is proud of the variety of fields outside the field of philosophy that he uses to support his theses. On the one hand, this is commendable because it shows the breadth of his interests, but in terms of the subject of research, the result is often negative, as examples and reasoning by analogy seem to further blur the issues instead of clarifying them. Thus, for example, when on p. 15 the author criticizes the thesis that the knowing subject has immediate access to his own knowledge, he refers to observations from neuroscience in relation to vision. It is obvious that we are not aware of the process of visual perception in all its complexity. However, what sensory visual processes have to do with the foundations of true discursive knowledge is not at all clear. Likewise, on p. 20, a finding about vision subtly becomes an argument for knowledge: "It is not the access to my visual perceptions, therefore, that is the basis of knowledge, but the reliability of the visual faculty or process." Philosophers have long spoken about the reliability of the visual ability. However, I am at a loss as to how this reliability of sight is the basis of the knowledge under epistemological analysis here with all its '-isms', justifications of justification, reliability, truth, etc.

The vocabulary and conceptual tools used in the dissertation are unnecessarily complicated. To justify the author, it can be pointed out that the language of the entire paradigm with which he works is such, and in this sense his possibilities for a different way of expression are limited. It is

obvious that orthodox Bulgarian epistemologists from this paradigm could hardly find in their native Bulgarian language equivalents of concepts and terms such as "reliabilism", "foundatinalism", "accessibilism", "fallibilism", "responsibilism", "mentalism" in this number and of the leading "internalism" and "externalism". One thing is certain, however, the drive to generate more and more "-isms" in the analysis and understanding of knowledge is more of a weakness than a strength. Other philosophical paradigms have also faced the problem of creating new means of expression. From Kant to Heidegger, understanding a particular philosophy goes hand in hand with understanding their specific vocabularies. In the tradition, however, this was not done lightly and self-servingly. And last but not least, it was not ruled by anyone. The case with the epistemology tradition from the end of the 20th and the beginning of the 21st century does not seem to be like that. The noisy debate between the hundreds and even thousands of internalists and externalists in recent decades has produced and disseminated this vocabulary strongly reminiscent of late scholasticism.

Philosophical platitudes are often hidden behind the pretentious terminology, which do not at all help to better understand the cognitive issues that Plato dealt with, with which the author of the dissertation also deals. For example, even in the introduction it is clear how externalism wins over internalism: "Culture, technology and the environment, more generally, have a decisive influence on how and what we think (sic). They are external to S factors. Therefore, externalism is preferable' (p.8). A few decades ago, the same trivial statement from the first sentence of the quote would have sounded similarly, albeit in a different ideological, but no less scholastic context, namely: "being determines consciousness."

Another example of how the author's complex thought processes lead to trivial conclusions is where the infinite regress in the rules of justification is considered and the author's response which boils down to this: "If I think that I am currently writing and at the moment write, then no further proof beyond my cognitive processes is necessary. When we speak of reliabilism, we shall mean this type of guarantee of belief' (pp. 65-6). With such a naive notion of what guarantees our knowledge, it seems superfluous to even ask the author whence and how he is sure that he is writing at the moment when he thinks he is writing? Is it possible that after abusing either philosophy or a hallucinogenic drug for example, his cognitive processes are so impaired that he thinks he is doing something he is not actually doing?

Another point where, instead of clarity, the concepts are blurred is the analysis of epistemic virtues, where the meaning of "virtue" in the moral and epistemic sense is entangled in analogies and associations. The author writes: "It is possible, like moral examples of good, to look for examples of epistemic good" (p. 47). Of course it is possible, but then the category "good" is not used in a moral sense and ceases to have anything to do with morality and ethics. In a moral sense, "virtue" cannot belong to knowledge per se, but only to the knower. However, the attempt to justify some rules of "good research" deserves admiration, which should be noted that this is not a particular novelty in philosophy. At least since Bacon, philosophy has been concerned with the rules that can guarantee not the virtue but the scientific standards of an inquiry.

I cannot fail to point out one misunderstanding in the dissertation. In his attempt to show the virtues of externalism, the author tries to show that logic could also rest on "external" empirical grounds. This is an extravagant move. The formal foundations of logic, according to the author, have two drawbacks: the first is that people are intuitively logical without having learned logic: “People, even babies, have an intuitive understanding of these relationships. This means that there are corresponding brain structures that perform the appropriate operations" (p. 50-51). Certainly humans have intuitions and brain structures that allow us to also be illogical and think nonsense, meaning the brain structures perform "inappropriate" operations. And in one case we have intuitions and brain operations, and in the other case we have the same. What follows from this observation is not at all clear. The second drawback is difficult to understand. According to the author, a disadvantage is that in formal logic "no objective or external verifier of belief is required" (p.51). Why should this be a disadvantage and not an advantage?

The subsequent reasoning confuses the matter even more, which necessitates a longer quote. Immediately after the sentence quoted above, the author continues as follows: “This goes against the foundations of logic, where it is a structural error to assert the consequent. If we had no external verifier to point us to some relationship between the antecedent and the consequent, then this belief would be justified. Let us say about the proposition: if it rains, it is wet; it is wet, therefore it is raining. It is a formal flaw, but this formality is derived from the observation (sic) that there may be other reasons for it to be wet. And these reasons are brought out of the mental state of the knower. This is a blow to internalism's desire to refer only to internal factors' (p.51).

This is indeed a blow, but in this case to the author himself. The logical error in this case is formal, i.e. it is due to the structure or form of the deductive inference, and in this case it does not

matter at all whether we are talking about the water in the streets, the sand in the Sahara, or any of the experience. To claim that a correct formal-logical inference on the one hand and the observation of wet streets on the other can have equal justification for belief is a clear misunderstanding.

The search for external validity of belief can be more clearly understood through the traditional category of objectivity. By other means, it seems that this is what the author is looking for when he reasons about the rules of justification that are of the "externalist type". This quest is authentic, but it also has many unclear moments.

According to the author, the rules of justification are external. In justifying the external nature of the rules, the author again resorts to sensibility, as if it were completely analogous to cognition. An error that was already pointed out. Here's part of the reasoning: "The main reason is that we don't choose the rules we follow. They are imposed on us by the outside world and will always find their basis in the outside world. We do not choose how the senses work; they are oriented towards and shaped by external factors. Only these external factors could, at the end of a study, show us to what extent they give justification" (p. 60-61). External factors define the rules, then confirm them. And here we can see the same logic circle that was mentioned above. The exterior of the rules, however, is not at all so certain. They can be no less "internal", i.e. created by the one who subsequently decides to follow them or not. Nor is there any reason to associate rules, whatever they are, with sensibility. When we create a certain methodology in science, for example, to ensure the reliability of the obtained knowledge, we do not take these rules from the outside world, nor from our sensibility. The same applies to the rules of morality.

When trying to answer the question of criteria for the validity of belief, the author uses the thesis' key category of "reliability": "The relation between truth and belief is characterized by a reliable cognitive mechanism that may be, but often is not, introspectively accessible." (p. 60). These mechanisms are central to the thesis, but clearly defining them proves extremely difficult. Exactly which processes are reliable and what is the criterion for reliability? Answers of the type that the reliability is guaranteed by some result remain unsatisfactory until criteria for the reliability of the results, etc., are in turn specified. to infinity. The author is aware of the regression problem and addresses it in a separate chapter of the study, but the result he reaches is far from satisfactory: "In most cases, our systems for knowing the world work adequately. In most cases, we orient ourselves in the world - what we see is what it is, what we think about something is what it is, etc.' when I write and my cognitive processes tell me that I am writing, then I am actually writing. Such

a naive and false conclusion effectively renders dozens of pages of examples and arguments in favor of reliabilism meaningless. For philosophers from at least the 5th century B.C. it is obvious that most things we think about are not what they are at all. If it were the other way around, as the author claims, we would never be dealing with epistemology and the questions of certainty, objectivity and scope of knowledge. Things would simply and plainly be exactly as they are and nothing more.

In the section where mental models are defined there is a curious moment in which the possibility of a "mental Darwinism" is vaguely sketched. It can be seen that after the attempts in the 20th century to justify social Darwinism, now at the beginning of the 21st century it is the turn of the intellectual Darwinism. According to this doctrine, "the model that survives is the one that best reflects the formal relations of the world" (p.81). How it is even possible to think of alternative "cognitive constructs", known traditionally as notions, ideas, theories, etc., as evolutionarily surviving thanks to the reflection of formal relations in the world is a complete mystery. Jordan Peterson's inspirational quote sounds like a mystical incantation - we could construct abstract models of "Being" and then have to let them die instead of us.

In relation to mental models and their relationship to culture and technology, I would like to ask the following question: Is it possible to somehow demonstrate and show through observation and experiment the "formation, functioning and expression" of a specific neural map of the world or with other words of a particular mental model? Take, for example, a simple model that any conscious person could have immediate conscious access to when uttering/reading the following expression "a cold beer on the beach". If the author's theses about mental models as neural maps of the world are correct, then he should be able to propose a methodology that identifies the specific characteristics of evolutionary pressures, genetic bases, features of prenatal and phylogenetic development, hormones, environmental influences, and finally the areas of activation in the brain that create the mental model corresponding to a cold beer bottle on the sand. If this is done, it will probably be possible to indicate in a similar way also what are the differences in cognitive processes responsible for the different neural maps of hot and cold beer and, accordingly, which of these models is evolutionarily more practical and has a higher chance to survive according to Intellectual Darwinism. How these questions can be answered under more complex neuro-mental models such as that of 'social justice' for example or 'aretic reliabilism' can be left for the author's future habilitation study, if the Philosophy department deems that should be invested in such an effort.

Abstract and contribution points

The dissertation abstract accurately presents the content of the work. Finally, a list of the contributing points, as well as the full bibliography to the dissertation, are presented as standard. Of the contributions listed, I disagree with the claim that a solution to the infinite regress problem has been proposed. I believe that the proposed solution is circular and even manipulative and does not actually solve anything. With the rest of the contributing points, I partially agree with the caveat that the presented theses are far from convincing.

I have no co-publications with the author of the study.

Conclusion

Despite the critical remarks and my disagreement with almost all the main theses and arguments of the author, I cannot help but notice that behind the dissertation text one can see a real effort to understand, an authentic interest in the subject and a wide awareness. I am convinced internally, consciously and responsibly, that the project of naturalizing philosophy probably needs even more young followers and admirers of Jordan Peterson, despite the fact that I myself deeply doubt the success of this philosophical program.

Based on all this and as a result of the above-mentioned inner confidence, I recommend the scientific jury to award Plamen Nikolaev Chergarov the educational and scientific degree "Doctor" in professional direction 2.3. Philosophy.

12.05.2024

Iassen Zahariev Zahariev

Digitally signed by Iassen Zahariev Zahariev
Date: 2024.05.12 12:58:30 +03'00'

Assoc. Prof. Iassen Zahariev, PhD,

New Bulgarian University