OPINION

on the dissertation of Plamen Chergarov

"Epistemological externalism in mental models"

by Assoc. Prof. Rosen Lutskanov, IPS-BAS

Plamen Chergarov's dissertation has a length of just over 200 standard pages and includes an introduction, five chapters divided into parts, a conclusion and a bibliography. Judging by the title, the study should be situated at the intersection of epistemology and philosophy of mind. Whether this is indeed the case, however, is difficult to judge, due to the many excursions justified by not always clear and well-founded reasoning by analogy, entering into other fields. At times this leads to various conceptual problems, which I will discuss below. In my opinion, I focus primarily on problematic points in the work that, in my view, indicate significant gaps in the PhD student's educational background and a failure to make use of a number of key conceptual distinctions.

The dissertation's task is to develop "a unified conceptual scheme that provides clarity regarding both the cognitive nature and mechanism of cognitive modeling of the world and the place of mental models in externalist doctrines" (p. 3). Accordingly, it’s goal is stated as "to defend epistemological externalism and its offshoots such as externalism about knowledge, externalism about justication, process and virtue reliabilism as the basic explanatory framework for mental models" (p. 6). It follows that the key concepts the PhD student is working with are 'epistemological externalism', 'mental model' and 'reliabilism'. As we shall see, there are problems in working with each of these concepts. In my view, these problems are partly due to the fact that no attempt has been made to explicate these concepts by proposing appropriate definitions and drawing appropriate distinctions on this basis. I would also point out that the weaknesses identified in the text are undoubtedly related to the difficulty and breadth of the chosen topic, which poses significant challenges.

Above all, it is imperative to distinguish externalism from its "other" - internalism. Unfortunately, the text offers neither a clear distinction between the two types of positions, nor an attempt to characterize them exhaustively, or at least to examine the systematics of the various theories within this field. I think that the main flaw in the author's approach is that it assumes that "internalism and externalism are not always mutually exclusive positions" (p. 13). The justification for this statement reduces to the claim that there are examples of justified true belief that satisfy both types of criteria. This begs the question of what it means for two positions to be "mutually exclusive." Option 1: two positions on X are mutually exclusive iff there is no x that belongs to X to which both attribute the same characteristic; Option 2: two positions on X are mutually exclusive iff taking one position on X precludes taking the other position on X. The above claim is justified via the first interpretation, but later in the text it seems that the second interpretation is adopted. Furthermore, we should bear in mind that, as mentioned in passing by the author, there are different types of internalism (internalism concerning knowledge and internalism concerning justification, access internalism and mentalist internalism), respectively different ways of denying them, which leads us to take externalist positions. The failure to distinguish between these options is evident, for example, on p. 20, where it is first stated that externalism 'regards conscious access as insufficient', and immediately thereafter that 'knowledge is possible even when its bearer is unable to indicate why or how it is possible'. These are entirely different claims, but the meaning differences between them are not thematized. The first claim presents externalism as a more demanding position than internalism, and the second as a more liberal, or "inclusivist" one (in this regard, see what is said on p. 13: "externalism simply makes a 'more inclusive' claim about the things that can be accepted as knowledge, whereas internalists have more stringent requirements").

Further, according to the way we have distinguished the two positions, part of the question comes down to which factors are internal and which are external. In this regard, mention is made of Ki-Hyeon Kim's position that "external is any element inaccessible to introspection" from which it naturally follows that "our cognitive activity ... is for the vast majority not internal" (p. 9). This raises the question of whether externalism is not decisively favored by this approach. We are led to the same thought by claims that "if mental models are influenced primarily by external factors ... then externalism is the preferred explanatory framework" (p. 9) and "insofar as knowledge is in relation to something outside the knower, it cannot be entirely internalist" (p. 47). In my view, this statement demonstrates a misunderstanding of what the two positions claim. I draw the same conclusion based on the claim that "[i]nternalism is right about knowledge being a kind of mental state, but the justifications for that state and the rules for their formation are of an externalist kind" (p. 50). Epistemological internalism is not a metaphysical thesis about what knowledge is; internalism about justification is a thesis about ... the justification of beliefs. Further, I would note that the question of 'what makes a belief justified' is quite distinct from the question of 'the justification for the justification of a belief', but the distinction is not drawn (cf. p. 44).

Finally, it seems to me that the text systematically fails to distinguish between descriptive claims (e.g., claims about how some type of cognitive process actually takes place) and normative claims (e.g., claims about what makes the products of such cognitive processes justified). Both internalism and internalism are normative theories of knowledge and justification. On the other hand, in a number of places descriptive generalizations are claimed to support them (in other words, an attempt is made to derive 'ought' from 'is'). For example, "unconscious information processing is supportive of the externalist project" (p. 95) and "the work defends the position that because we do not understand the rules we follow to know the world, but they are determined by the world itself, justificatory factors are primarily externalist" (p. 189). In conclusion, I would note another odd characteristic of the text. It identifies "the problem of infinite regress" as "one of the serious criticisms of the externalist project" (p. 63; see also p. 189). It is not clear to me why this should be a problem for externalism specifically and not for internalism (or, better yet, for epistemology in general).

I will now move on to the concept of mental model, which I also find inadequately employed. First, the author declares that an external model becomes a mental model when it is used by a cognizant subject (p. 73). On the other hand, we are offered the following 'definition': 'I argue that if we define a mental model as a model that enables interaction with the world by ordering the information that comes in a meaningful way, then the information taken in is part of the mental model even though it is external to it' (p. 79). The understandings of mental model expressed in the two statements are mutually orthogonal: if in the first case something becomes a model because it is used in a certain way, in the second it is a model because it makes a certain type of information processing possible (it is puzzling, by the way, how something that is part of a whole can at the same time be external to it). Another problematic point is that the notion of a mental model as used in the thesis removes a number of key distinctions: for example, it does not distinguish between either a mental model and a representation, or between a mental model and a cognitive schema (on the occasion of the latter, a mental model is defined as an 'epistemic faculty', further complicating the interpretation of what is said) (p. 84). Furthermore, the notion of model is stretched (almost) to breaking point by explaining to us that universal statements are also models (p. 82). Accordingly, the extension of the notion of model leads to an extension of the notion of mental model, which is clearly a problem for the present work. Further, in relation to the notion of model, the author introduces four theses: (1) "a mental model is a representation" (although we have just seen that cognitive schemata are also conceived as mental models); (2) "it is composed of various perceptual and/or semantic elements" (which in turn leads us to ask: if perception gives rise to sensory representations, which are mental models, then how are mental models composed of perceptual elements); (3) "it creates a map of the world that allows interaction with the world itself" (this echoes what p. 79 claims and abandons what was said on p. 73); (4) "the relation between mental model and neural structure is isomorphic" (In this regard, I would like to emphasize the following: (i) nowhere in the text is there any attempt to explain what "isomorphic" means in this case, in other words, between which components of the mental model and of the neural structure there is a 1-to-1 correspondence; (ii) nowhere in the text is there any attempt to substantiate this claim; (iii) this thesis is denied before and after this passage by telling us that thoughts are "identical to the activity of neural networks" (p. 83, 190) and that the work accomplishes an "equating" or "reduction" of mental to neural pattern) (pp. 85-86). In this connection, I want to emphasize the following: (1) the claim that something is identical to something else; (2) the claim that something can be equated to something else; (3) the claim that something can be reduced to something else; and (4) the claim that something is isomorphic to something else are different claims. The smooth slippage between them is symptomatic of the imprecise handling of concepts that is a problem for the work in its entirety.

Before I go any further, I will provide the following "important insight" that connects the topic of mental models to the topic of externalism, which is the main purpose of the paper: "Assuming that they [mental models] are maps of the world that are isomorphic to neural networks, we conclude that the models are driven primarily by automatic processes. These processes must work reliably for us to have knowledge, and this reliability is determined by the interaction with the environment. Because the environment has enduring and demonstrable influences on the operation of these mechanisms, we can assume that externalism has greater explanatory value than internalism with respect to the cognitive value of mental models" (pp. 101-102). This passage cashes in on all of the conceptual conflations and shifts that we have identified above: (1) an inference is drawn from isomorphism that can only be justified by identity; (2) a normative claim about justification is defended by a descriptive claim about the influence of the environment on cognitive mechanisms; (3) what exactly is being explained in this case is not explained, making it impossible to judge whether the claim about the explanatory value of externalism is successfully defended.

The third key concept, 'reliabilism', has not received much attention. However, what has been said is sufficient to justify some conclusions about dealing with this concept. I find puzzling the claim that 'reliabilism is part of an virtue epistemology' (p. 162), and thus the talk of 'reliabilism and other virtue approaches' (p. 163). Reliabilism is not part of virtue epistemology; as is well known according to Goldman, it is correct to define them rather as "cousins". It is true that there is such a thing as virtue reliabilism which, according to its name, belongs to both currents in epistemology. It connects to the topic of the dissertation because, according to him, mental models can be treated as "integrative structures for cognition" (pp. 30-31), respectively as the basis of epistemic virtues. Exactly how is not entirely clear. The justification for this claim comes down to the example of John von Neumann, which is presented in an entertaining way to say the least: 'The models he used were drawn from heavy technical textbooks. What Neumann did was to turn these textbooks to the last page that contained the formulas and learn them by heart" (p. 183). In my view, there is nothing particularly virtuous about memorizing formulas, and I am extremely skeptical that by turning textbooks to the last page you can develop an epistemic competence comparable to von Neumann's. Of course, I'm only expressing my personal opinion here.

I will conclude with a couple of isolated problematic points that I find difficult to pass over without comment: (1) The explanation of the meaning of "supervene" on p. 53 is inaccurate: X supervenes on Y not when it "is based on or a consequence of it but distinct from it," but when it is impossible to have differences in X without having differences in Y; (2) the idea that "a person's conviction can be measured by his willingness to bet a certain amount on a given outcome" (p. 66) certainly does not belong to the mentioned author, how old exactly is it is debatable; (3) the attempt to justify the role of externalism in terms of mental models by examples related to critiques of CRT, BLM, LGBT, and the like (see ch. 4) is, in my view, utterly unsuccessful. The presentation of the debate here amounts to a reproduction of extremely simplistic and stereotyped ideas on the matter; (4) the claim that "Kant insists that we have embedded models that determine our perception of the world" (p. 134) further expands the already stretched notion of mental model and is anachronistic in the extreme. Kant speaks of a priori forms, not of "embedded patterns"; (5) reducing Ebbinghaus's contributions to an awareness of the fact that "the more one repeats a word, the easier it is to learn" and that "remembering 12 words at once is harder than remembering 6" (p. 149) is grossly unfair. The forgetting curve and the introduction of experimental methods in the study of memory are what the history of science has remembered this author for; (6) The bibliography credits the company Amazon with Sapolsky's course “Biology and Human Behavior” (p. 196). Overall, the bibliography is extremely rich and includes many interesting titles, but many of these are popular in nature and have no direct bearing on the main topic of the text.

Finally, I would like to comment on the stated contributions. I did not notice a "solution to the problem of infinite regress" in the text. I don't think that the justification of the thesis that culture, hence technology, "has an enormous influence on mental models and therefore their explanation must be within an externalist framework" can be considered a contribution (again, it is not clear to me what exactly is being explained and how explanatory properties are ascribed to a normative theory).

May 19, 2024,

Sofia, /R. Lutskanov/