Towards Model-Based Model of Cognition

In data mining, people are working (sometimes - with almost no theory behind) with a variety of computer models that are extremely useful in practice, but hardly resemble their target objects. "All models are wrong, but some are useful." - as put by the statistician [George E. P. Box](#) (1979: Robustness in the strategy of scientific model building. *Robustness in statistics,* 201-236).

This wide experience with "theory-less, wrong and useful" models is provoking three radical ideas. None of them is completely new, the novelty might be the composition proposed below: let's consider philosophy of all kinds of cognition, and try a completely model-based rewriting of it.

The first idea: *internally,* there is no difference between "true" models and "wrong" models. Mainly, we do not know in advance (sometimes - for centuries), which parts of our models or theories are true, and which are not. Sometimes we know that our model is "wrong" in many respects, but it remains useful, nevertheless. Many people may continue believing in an overthrown model or theory for a long time - and continue acting accordingly. Hence, the somewhat paradoxical term - "wrong" knowledge. This corresponds well to "The Dappled World" picture ([Nancy Cartwright](#)), see [Paul Teller](#) (2004: [How We Dapple the World](#), *Philosophy of Science*, 71(4): 425-447). Thus, for the philosophy of cognition, *"wrong" knowledge should be as prominent a subject as the "true" knowledge.*

And, if we wish to answer non-trivial questions like as "Is it true that quarks really exist?", then our *philosophy of cognition shouldn't introduce the notion of "truth" too early,* as something primary, and therefore - mystical. Truth is emerging later - on top of cognition, it doesn't reside at the bottom of it. We can derive ontologies and truths only by analyzing *invariants* of a successful model evolution - as I tried to propose in "Is Scientific Modeling an Indirect Methodology?" (*The Reasoner* 3(1): 4-5).

In terms of theories, a similar idea was proposed as the "Deepening Maxim" by [Paul Thagard](#) (2007: [Coherence, truth, and the development of scientific knowledge](#). *Philosophy of Science*,74, 28-47): "If a theory not only maximizes explanatory coherence, but also broadens its evidence base over time and is deepened by explanations of why the theory's proposed mechanism works, then we can reasonably conclude that the theory is at least approximately true." (41). And by [Jeffrey Alan Barrett](#) (2008: Approximate Truth and Descriptive Nesting. *Erkenntnis* 68(2): 213-224) - as "a notion of local probable approximate truth in terms of descriptive nesting relations between current and subsequent theories" (213).

The second idea: models are the *ultimate results* of all (scientific, non-scientific, and anti-scientific) kinds of cognition. Therefore, philosophy of cognition should start with the following fundamental distinction: *there are models, and there are means of model-building.* Laws of nature and theories are useful only as a means of model-building - "The Toolbox of Science" - as put by [Mauricio Suarez](#), Nancy Cartwright (2008: [Theories: Tools versus Models](#). *Studies in History and Philosophy of Modern Physics*, 39: 62-81).

The third idea: to cover all kinds of cognition, the notion of model should be defined as broadly as possible: *a model is anything that is (or could be) used, for some purpose, in place of something else.* To put it somewhat paradoxically: models are tiny fragments of the Universe possibly usable (for some purpose) in place of other fragments (or, even in place of the entire Universe). Mathematical models, fictional worlds, mental structures and processes are included here, of course.

Among philosophers, this broadest possible notion of model was stated by Paul Teller (2001: Twilight of the Perfect Model Model. *Erkenntnis*, 55: 393–415). But, among computer scientists, it can be traced back to [Marvin Minsky](#) (1965: [Matter, Mind and Models](#). *Proceedings of IFIP Congress 65*, 1: 45-49).

Minsky applies the notion of model in a way, that is very natural from computer scientist's point of view, but seems not very popular among philosophers. I would put this "robotic ontology" as follows:

In my head, I have a model of the world (an incomplete one, incoherent, inconsistent, in part - fictional, containing all my knowledge, beliefs, dreams etc.). And I'm acting according to this model. In this model, other persons are believed to have their own models of the world (in some respects - different from my model). And they are acting according to their models. I may know these models more or less, and in this way I can predict - to some extent - people's behavior. Thus, my model of the world may contain "models of models" - for example, a simplified model of your model of the world.

And, to complete the picture: how about model-building in philosophy? Perhaps, many will agree with Peter Godfrey Smith (2006: Theories and Models in Metaphysics. *The Harvard Review of Philosophy*, XIV: 4-19): "... much metaphysical work, especially of the contemporary systematic kind, might best be understood as model-building, ..." (4). However, "It would be foolish to suppose that such a hypothesis could be applied to all metaphysical discussion, but it might be true of an important part of the field." (5). But how about trying this "foolish" step?

If none of the above theses is completely new, then - what is missing? My general impression: despite many brilliant insights, generated by philosophers for many years, the field (philosophy of cognition) remains unordered for too long a time. For example, according to the account given by Roman Frigg, Stephan Hartmann (2006: Models in Science. *Stanford Encyclopedia of Philosophy*), there is still no generally acknowledged unified notion of model. Or, according to the account given by Eric Schwitzgebel (2006: Belief. *Stanford Encyclopedia of Philosophy*), there are several competing approaches to explaining "what is it to believe". Couldn't these complications be caused by the idea of "propositional attitude" - the idea that separate propositions are believed in, and not entire models, theories, or fragments of them?

If it's true that models are the ultimate results of cognition, then shouldn't we try reordering the field, starting with the notion of model? In this way, couldn't we obtain a unified and more productive picture - a model-based model of cognition?

Karlis Podnieks
Computer Science, University of Latvia