Explanation and Understanding in a Model-Based Model of Cognition

Karlis Podnieks

University of Latvia

Preprint: version November 20, 2017

Abstract. This article is an experiment. Consider a minimalist model of cognition (models, means of model-building and history of their evolution). In this model, explanation could be defined as a means allowing to advance: *production of models and means of model-building* (thus, yielding 1st class understanding)*, exploration and use* of them *(*2nd class), and/or *teaching* (3rd class). At minimum, 3rd class understanding is necessary for an explanation to be respected.

This article is an experiment. Imagine a minimalist picture of cognition (model-based model of cognition, MBMC) where we and our robots have *only* models, means of model-building and history of their evolution. In this picture, the most significant distinction exists between *particular models* (mainly, serving as *replacements* of concrete target systems) and *means of model-building* (model templates, theories, methods, hypotheses, heuristics, research programs, doctrines, paradigms, frameworks, ontologies, metamodels, metametamodels, mathematical structures, logic systems, languages, etc.).

I have been trying to promote MBMC since publishing Podnieks (2009). The ideas on which MBMC is based were proposed separately by several authors. For an account of the corresponding history, see Podnieks (2017).

How to recognize the phenomenon of explanation and understanding (subtleties aside) in the picture of “only models, means of model-building and history of their evolution”? First, let us look for further significant distinctions in this picture.

There are means of model-building (for instance, the causal paradigm) that are used as a guide for building of theories, so, they serve as *meta-means* of model-building. The *meta-level* aspect represents a significant distinction in the picture proposed by MBMC.

The *evolution* aspect is significant as well: more or less long periods of stability, more or less radical changes. Some of the models and means of model-building are used for limited time periods only (such as phlogiston theory of combustion, or plum pudding model template of atoms).

Two other significant aspects: the *user community* aspect and the *application domain* aspect. Models and means of model-building are accepted and used by more or less wide communities. And, they are applied across more or less wide domains of knowledge. For instance, the traditional realist idea (in the narrow sense – the idea that my sensations are caused by an independent "reality" populated by creatures similar to me) is extremely widely accepted and applied. The causal paradigm is widely accepted and applied as a guide as well, but it has some limitations.

What should be counted as more important: *observable* significant distinctions (like as the above ones), or their correspondence to intuitive notions such as truth and explanation? If we wish, we can try establishing of such correspondences.

For example, we can try to introduce a concept of *truth*. MBMC inspires a kind of pragmatic-operational definition of truth: truths are more or less *persistent invariants of successful evolution* of models and means of model-building (evolution aspect). What is true will not change in the future (for some time, and for some of us, at least – user community aspect). The traditional realist idea represents the most fundamental invariant of model-building, hence, the first truth to believe in.

How about explanation and understanding? For inspiration, let us start with the problem of understanding in mathematics. As William Thurston (1994) put it:

"... when Appel and Haken completed a proof of the 4-color map theorem using a massive automatic computation, it evoked much controversy. I interpret the controversy as having little to do with doubt people had as to the veracity of the theorem or the correctness of the proof. Rather, it reflected a continuing desire for *human understanding* of a proof, in addition to knowledge that the theorem is true." (p. 162).

"Finally and perhaps most importantly, a mathematical breakthrough usually represents a new way of thinking, and effective ways of thinking can usually be applied in more than one situation." (p. 172)

"More than the knowledge, people want *personal understanding*." (p. 173)

"... what was dramatically lacking in the beginning: a working understanding of the concepts and the infrastructure that are natural for this subject [geometrization conjecture]." (p. 175)

Thus, according to Thurston, a mathematical idea leads to understanding, if it proposes a "breakthrough," "new way of thinking," an "infrastructure" allowing for "more than one" problem of the field to be solved. Notably, computer-generated parts of mathematical proofs, as a rule, do not yield far-reaching infrastructures. Hence, the controversy around such proofs.

Note. As the next step, one could try connecting Thurston's argument to the famous Rising Sea strategy promoted by Alexander Grothendieck, see McLarty (2007).

Let us apply Thurston's argument to the whole of cognition. Then, in terms of MBMC, we obtain the following thesis: more than particular good predictive and action-recommending models, and specific means of model-building people want stable (evolution aspect) and widely applicable (application domain aspect) means allowing to produce, explore, use and/or teach models and means of model-building.

What is added to our knowledge by an act of explanation? It seems, in MBMC, we could define explanation as a means allowing to advance: *production of models and means of model-building* (thus, yielding 1st class understanding)*, exploration and use* of them *(*2nd class), and/or *teaching* of them (3rd class). Having understood X, I can build better models for the world around X. Or, at least, I can teach them better: at minimum, 3rd class understanding is necessary for an explanation to be respected.

For example, Kepler's introduction of elliptic orbits (instead of epicycles used by Ptolemy and Copernicus) was a great act of 1st class understanding that allowed not only for building of radically simpler models of the Solar system, but also contributed to the 1st class idea of gravitation and all the great development that followed it.

From this perspective, special relativity and quantum mechanics represent, despite their counter-intuitiveness, the greatest acts of 1st class understanding achieved in the 20th century. But what about *interpretations* of quantum mechanics – should we qualify them as merely 3rd class?

Many people writing about explanation pursue the following *strategy A*: try to identify some significant distinction in the human cognition which could be verified as being close enough to the "well known" intuitive notion of explanation. As a rule, they succeed in the identification of the distinction, but (as noted later by the numerous opponents) do not succeed in approaching closely enough the intuitive notion. My above proposal could be rejected on similar grounds. Therefore, I would propose to revert strategy A, and replace it by *strategy B*: identify a significant distinction in the human cognition, name it as XYZ, and prove that XYZ is more significant than the possibly corresponding intuition. Which of the two strategies will appear more productive?

References

Colin McLarty (2007): The Rising Sea: Grothendieck on simplicity and generality. In: Jeremy Gray and Karen Parshall (Eds.), *Episodes in the History of Modern Algebra (1800-1950)*, Amer. Math. Soc., 301-326.

Karlis Podnieks (2009): Towards a Model-Based Model of Cognition. *The Reasoner,* 3(6), 4-5.

Karlis Podnieks (2017): Philosophy of Modeling: Some Neglected Pages of History, *PhilSci Archive* 13538 [preprint].

William P. Thurston (1994): On proof and progress in mathematics. *Bull. Amer. Math. Soc.*, 30(2), 161-177.