Logical and epistemological criteria of salience in the foundations of mathematics

The main objective of our project is to provide a better understanding of the asymmetry between *salient* (or *natural*) and *artificial* (or *ad hoc*) theories in the foundations of mathematics. The objects falling under our scrutiny, i.e. formal systems, are well-defined mathematical entities but the above mentioned salience criterion is pre-theoretical. Thus, our project falls within the scope of *formal philosophy*, in which mathematical and philosophical investigations are tightly intertwined.

While from a formal perspective any set of sentences can be called a theory, intuitively each theory should be a *theory of something*: it should describe a real world phenomenon, capture the behaviour of a certain concept or at least fulfill one of these tasks to a certain reasonably high degree. Many theories studied in the foundations of mathematics have this property: they derive from our analysis of the structure of natural numbers (such as the canonical theory of natural numbers: Peano Arithmetic, PA), results in analysis (subsystems of Second Order Arithmetic), investigations into the concept of truth (Axiomatic Theories of Truth) or the hierarchy of sets (such as Zermelo-Fraenkel set theory, ZF). These are often called *salient* or *natural* theories and their distinctive feature is that they "have something like an 'idea' to them".

Is any theory akin to the ones mentioned above? Let's concentrate on theories in classical first-order logic that are conjectured to be consistent. There is a sense in which an arbitrary such theory is a theory of something: for each consistent theory there is a rigourously defined mathematical structure in which all the axioms of this theory are true. In spite of this, there are many theories which seem to lack any kind of deeper foundational motivation and appear to be just artificially created, consistent collections of sentences. Most often they occur as counterexamples to some simple-minded, but typically intuitive, generalisations based on experiences from the world of natural theories. For example, the above list of natural theories suggests the following conjecture: every natural theory U has an immediate successor, i.e. a natural theory V which is stronger than U and such that no theory is strictly in between U and V. So, is it true that *every* theory has an immediate successor in this sense? Emphatically *no*: one can show (for example) that there is no logically weakest proper extension of PA (hence creating infinitely many different theories whose "infimum" is PA). Even more shockingly, there is a family of extensions of PA such that the provability relation between them reflects the ordering of rational numbers. This picture reemerges for every salient theory considered above.

With the examples introduced in the previous paragraph in mind we can ask whether there is any deeper, conceptual difference between *natural* and *artificial* theories? Is it just an accidental phenomenon that some formal systems attracted theorists' attention or do they owe their salience to some deeper intrinsic feature? Are salient theories separable from the artificial ones by a formal criterion? These are the questions that motivate the research within this project.

In the project the distinction between salient and artificial theories is approximated from three different angles. Firstly, we scrutinize the intuitive view that natural theories arise from comprehensive descriptions of salient mathematical structures, such as the natural numbers or fragments of the hierarchy of sets. We shall investigate, both from philosophical and mathematical perspectives, properties of first-order theories such as *tightness*, *solidity*, and *internal categoricity*. We aim at verifying whether any of these formal notions provide us with a good explication of salience in the case of foundational theories. If this is so, we shall gain an extremely intuitive formal notion that will coincide with a pre-theoretical concept of a salient theory. This will be one of, unfortunately, few cases in which our philosophical intuitions find a precise formal explication. Secondly, we analyze a philosophical property of formal systems, known as *epistemic stability*. Roughly speaking, an epistemically stable theory *captures* a standpoint in the philosophy of mathematics. By this we mean, that the theory is sound and complete with respect to the methods of developing mathematics according to the standpoint. Two most prominent examples of systems which exemplify this property are: *Primitive Recursive Arithmetic*, which catpures finitism and *Ramified Analysis up to* Γ_0 which captures predicativism. We concentrate, among others, on determining which properties of formal theories are necessary for their epistemic stability. Finally, the third dimension of our project consists in seeking coordinate-free characterizations of formal systems, that is complete descriptions of the given system that mention neither the language in which the system is formulated, nor its specific axiomatization. Coordinate-free characterizations abstracts away some, arguably, more accidental features of formal systems and uncover their more intrinsic properties. The background idea is that each choice of a language may give us a different perspective on a foundational idea much in the same way various choices of coordinates give us different perspectives on a moving object in physics. However, the idea itself is mostly independent of its various linguistic descriptions. The common intuition among logicians is that salient theories shall admit such coordinate-free characterizations, however we are still quite far from realizing this goal. Our project is a step in improving this situation.