The Reflective Equilibrium of Intended Models

Word count: 1395

Abstract

The categoricity theorem shows that the realist is justified in believing that the sentences of a foundational theory $\mathcal{T}$ – such as Peano arithmetic, Cantor-Dedekind analysis and Zermelo-Fraenkel set theory – have a determinate truth-value if she is independently justified in believing that $\mathcal{T}$ has a particular model. In the first part of my talk, I will point out an overlooked connection between categoricity and antecedent beliefs, namely that of *extremal axioms* – such as the axioms of Induction, Continuity and either Constructibility or Large Cardinals. Extremal axioms define a condition of either *maximality* or *minimality* on the class of models satisfying $\mathcal{T}$. Moroever, given some additional constraints, the assumption of extremal axioms implies the categoricity of $\mathcal{T}$. In the second part, I will argue that the assumption of extremal axioms is justified by the *reflective equilibrium* between the antecedent beliefs and the formal resources adopted to formalize the theory. I will claim that the proposed framework meets the epistemic *desiderata* imposed by the reflective equilibrium method. In the last part, I will consider two case studies – namely, that of arithmetic and set theory – showing that the assumption of extremal axioms is adopted to revise, respectively, the formal resources and the antecedent beliefs.

Introduction

Mathematical theories might be distinguished into those that are about a specific subject matter (called *foundational* theories) – such as Peano arithmetic, Cantor-Dedekind analysis and, possibly, Zermelo-Fraenkel set theory – and those that are instead designed for different mathematical branches (called *algebraic* theories) – like algebraic structures and topological spaces. Mathematicians and philosophers generally agree that foundational theories have an intended model or interpretation – respectively, the natural numbers, the arithmetic continuum and, possibly, the set theoretic universe. That is why, the (quasi-)categoricity theorem is usually considered as a desirable property of foundational theories. More precisely, a Second-order theory $\mathcal{T}$ characterizes the (intended) model $\mathfrak{M}$ *iff* two conditions hold:

$$\mathfrak{M} \vDash \mathcal{T} \qquad \text{(Existence)}$$

$$\forall \mathfrak{M} \forall \mathfrak{N} ((\mathfrak{M} \vDash \mathcal{T} \wedge \mathfrak{N} \vDash \mathcal{T}) \rightarrow \mathfrak{M} \cong \mathfrak{N}) \qquad \text{(Uniqueness)}$$

However, several authors in the literature have warned from overestimating the philosophical relevance of the categoricity theorem (i.e. the uniqueness thesis) – see Walmsley (2002) and Meadows (2013). Categoricity shows that the realist is justified in believing that the sentences of the foundational theory $\mathcal{T}$ have a determinate truth-value if she is independently justified in believing that $\mathcal{T}$ has a particular model (i.e. the existence thesis). Several proposals have been made in the literature to establish the existence of such models, depending on the background resources of either Second-order Logic or Zermelo-Fraenkel set theory.

In this talk, I will propose a coherentist framework according to which the belief in the conditions of existence and uniqueness is revised according to the assumption of *extremal axioms*. More precisely, the assumption of extremal axioms is made to "test" the belief in the existence thesis against that of the uniqueness thesis, in order to establish an agreement between the two. Extremal axioms are a set of principles which comprehends the axioms of Induction in Peano arithmetic, the axiom of Continuity in Dedekind-Cantor analysis, and either the axiom of Constructibility or Large Cardinals in Zermelo-Fraenkel set theory. Extremal axioms determine a condition of either minimality or maximality on the class of models $\mathfrak{C} := \{\mathfrak{M} : \mathfrak{M} \models \mathcal{T}\}$ defined by the theory $\mathcal{T}$. In this sense, extremal axioms specify the size of the model $\mathfrak{M}$ postulated by the existence thesis. More precisely, model minimality (maximality) means that each model in the class of models admits no proper restriction (extension) to a model satisfying the theory, namely:

$$Min(\mathfrak{C}) := \{\mathfrak{M} \in \mathfrak{C} | \forall \mathfrak{N} \in \mathfrak{C} : \mathfrak{N} \subseteq \mathfrak{M} \rightarrow \mathfrak{N} = \mathfrak{M}\}$$

$$Max(\mathfrak{C}) := \{\mathfrak{M} \in \mathfrak{C} | \forall \mathfrak{N} \in \mathfrak{C} : \mathfrak{M} \subseteq \mathfrak{N} \rightarrow \mathfrak{N} = \mathfrak{M}\}$$

Moreover, given some additional constraints, the assumption of extremal axioms entails the categoricity of the theory – thus granting the uniqueness thesis, see Carnap et al. (1936). For instance, as the axiom of Induction implies that the natural numbers are the *minimal* set closed under the successor function, so the axiom of Continuity implies that the real numbers are the *maximal* Archimedean ordered field.

I will argue that the assumption of extremal axioms is justified by the *reflective equilibrium* between the antecedent beliefs and the formal resources adopted to formalize the theory. Reflective equilibrium (RE) is a coherentist account of epistemic justification tracing back to the work of Goodman (1952), according to which knowledge of some specific domain is reached through a process of mutual adjustment among particular judgements and general principles. More precisely, let ϕ be the First-order extremal axiom of Induction/Continuity/etc. and let Φ be its Second-order formulation. Then, the coherence of our judgements about the intended models of foundational theories is obtained trough the RE between the belief that *'It is true that ϕ'* and the belief that *'There exists a class of models $\mathfrak{C}$ satisfying $\mathcal{T}_\phi$'*. Clearly, the two beliefs correspond to, respectively, the judgment and principle of the RE method – more on this below. The RE between the judgement and principle is reached once $\mathcal{T}_\Phi$ turns out to be categorical. For example, someone might start by judging the axiom of Induction as a 'self-evident truth'. Having considered the existence of non-standard models satisfying the First-order theory of Peano arithmetic, she revises the formalization of the Induction axiom by means of Second-order logic so as to achieve a categorical theory. Then, the categoricity result further supports the initial belief in the existence of a minimal model of the natural numbers. I will support the proposed framework arguing that the judgement of ϕ and the principle concerning the class of models $\mathfrak{C}$ satisfying $\mathcal{T}_\phi$ meet the epistemic *desiderata* imposed by the reflective equilibrium method. Indeed, while judgements should posses an epistemic standing of initial credibility, principles should achieve some epistemic goals which motivate the transition from judgements in the first place – see Daniels (2020). Finally, I will consider a possible objection concerning the background theory of Second-order logic and its full semantics adopted to establish the categoricity result.

In the last part of my talk, I will apply the proposed framework to the case studies of arithmetic and set theory. I will point out that the mutual adjustments between judgement and principle highlight the different philosophical lessons drawn from the construction of non-standard models. While the existence of non-standard models satisfying Peano arithmetic motivates the revision of the formalization of the theory, the forcing extensions of the set theoretic universe led to revise the antecedent beliefs concerning the subject matter of the theory. More precisely, in the case of arithmetic, the existence of non-standard models prompts for the Second-order formulation of Peano arithmetic. As mentioned above, the resulting categorical theory further supports the initial plausibility of the extremal axiom of Induction, which characterizes the natural numbers as the minimal set closed under the successor function. Instead, in light of the quasi-categoricity of Zermelo-Fraenkel set theory, Godel started the search for maximality (and minimality) principles which would settle independent sentences such as the Continuum Hypothesis. For instance, while Godel initially believed that the set theoretic universe is determined by a minimal property (i.e. the axiom of Constructibility), he came later to realize that only a maximal property (i.e. the axioms of Large Cardinals) would work. Replacing the extremal axiom of a theory corresponds to revise the belief in the existence thesis, i.e. the size of the model satisfying the theory. However, the tension between the extremal axioms of Zermelo-Fraenkel set theory and the forcing extensions of its Second-order models have cast doubts over the full categoricity of the set theoretic universe – opening up to the multiverse position in the philosophy of set theory, see Hamkins and Solberg (2020). Therefore, the failure of establishing the full categoricity of $\mathcal{T}_\Phi$ ultimately led to reject the uniqueness thesis. I will conclude that the method of RE vindicates the realist's belief in the determinacy of arithmetical sentences, regarding instead axiomatic set theory as an algebraic theory.

References

Carnap, R., Bachmann, F., and Bohnert, H. G. (1936). On extremal axioms. *History and Philosophy of Logic*, 2(1-2):67–85.

Daniels, N. (2020). Reflective Equilibrium. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2020 edition.

Goodman, N. (1952). Sense and certainty. *The Philosophical Review*, pages 160–167.

Hamkins, J. D. and Solberg, H. R. (2020). Categorical large cardinals and the tension between categoricity and set-theoretic reflection.

Meadows, T. (2013). What can a categoricity theorem tell us? *The review of symbolic logic*, 6(3):524–544.

Walmsley, J. (2002). Categoricity and indefinite extensibility. In *Proceedings of the Aristotelian Society*, volume 102, pages 239–257. Oxford University Press Oxford, UK.