

On representations of intended structures in foundational theories

Neil Barton*, Moritz Müller[†] and Mihai Prunescu[‡]

04 April 2021[§]

Abstract

Often philosophers, logicians, and mathematicians employ a notion of *intended structure* when talking about a branch of mathematics. In addition, we know that there are *foundational* mathematical theories that can find representatives for the objects of informal mathematics. In this paper, we examine how faithfully foundational theories can represent intended structures, and show that this question is closely linked to the decidability of the theory of the intended structure. We argue that this sheds light on the trade-off between expressive power and meta-theoretic properties when comparing first-order and second-order logic.

Keywords: structure; representation; first-order logic; higher-order logic; decidability;

*Fachbereich Philosophie, University of Konstanz. E-mail: neil.barton@uni-konstanz.de.

[†]Universitat Politècnica de Catalunya. E-mail: moritz@cs.upc.edu.

[‡]University of Bucharest, Faculty of Mathematics and Informatics; and Simion Stoilow Institute of Mathematics of the Romanian Academy, Research unit 5, P. O. Box 1-764, RO-014700 Bucharest, Romania. E-mail: mihai.prunescu@imar.ro, mihai.prunescu@gmail.com

[§]Order of authors is alphabetical. We would like to thank two anonymous reviewers for very helpful comments on the paper. The first author is grateful to the VolkswagenStiftung for their support through the project *Forcing: Conceptual Change in the Foundations of Mathematics*. The second author is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement ERC-2014-CoG 648276 (AUTAR).

Introduction

This paper addresses the philosophical question of how well foundational mathematical theories are able to represent mathematical structures. Much of mathematical practice concerns the study of particular structures. Famous examples are the arithmetical structure of the natural numbers $(\mathbb{N}, +, \times, 0, 1, <)$, the ordered field of the reals $(\mathbb{R}, +, \times, 0, 1, <)$, or the field of complex numbers $(\mathbb{C}, +, \times, 0, 1)$. The study of these structures is conducted mainly informally, such as the manner of reasoning we see in mathematical journals.

This fact concerning mathematical practice is coupled with the existence of *foundational* theories. There are various such theories, first and foremost set theory *ZFC*, but also category theory, and more recently homotopy type theory. There are many features we might want a foundational theory to have, but two (interlinked) desiderata that have emerged are: (1.) to provide a *generous arena* for mathematical discourse—we want to provide proxies for all the objects of informal mathematics, and (2.) to yield a *shared standard*—we want to be able to codify informal proofs in the theory in order to compare them and say when a construction or proof counts as legitimate.¹ Roughly, this means that the foundational theory can encode or formalise all our informal mathematical discourse about the ‘usual’ objects of mathematics. In this way, if one had sufficient patience and time, one could formalise all theorems of informal mathematics as theorems within one’s favourite foundational theory. The starting question of this short paper is: What is the desired relationship between informal and formalised mathematics?

Being a very general question we restrict attention to the informal study of concrete structures, like the natural, real or complex numbers mentioned above; in a philosophical context these are often referred to as *intended* structures. Now, it is one thing to be able to formalise some piece of informal mathematics any-old-how, and quite another to do so faithfully. We would like the intuitive meaning of the formal statements to be similar to the intuitive meaning of the informal statements.

For motivational purposes let us roughly distinguish two approaches to the foundations of mathematics: the *axiomatic* and the *genetic* method (see [Rav, 2008]).² The first, chiefly embodied by Hilbert, replaces the intended structure by a set of axioms we argue are (or take

¹We refer to Maddy [Maddy, 2019] for an assessment of what we want a foundational theory to do for us. The terms ‘generous arena’ and ‘shared standard’ are her terms.

²A similar distinction is often referred to in mathematics texts as the difference

to be) true. When done in first-order logic this approach is often incomplete (by Gödel's results). When done in higher-order logics, we lose various pleasant meta-theoretic properties, and so whilst of philosophical interest it has less practical value.

The genetic approach is via construction. Instead of asserting axioms for the intended structure, one first constructs the structure in question by finding an object coding it in one's foundational theory, and then one asks about its properties. Theories with a high degree of interpretive power are able to translate some mathematical constructions into first-order definitions inside the theory. For example, ZFC can mimic the classical constructions of the natural, real or complex numbers by defining formulas. In a little more detail (we provide a full outline in §2) formalisation of the mathematical study of an intended structure \mathcal{S} typically proceeds in two steps. First, \mathcal{S} is represented by a sequence of formulas R that identifies an object within the foundational theory (in the case of ZFC, a set or a class). Second, the informal talk about \mathcal{S} is translated to formal talk about R . Thus, in every model of ZFC we will find an avatar of the natural, real or complex numbers.

Here, we are concerned with the first step; the choice of R .³ We will focus mainly on this style of doing mathematics: by first constructing the structures and then examining their properties, and especially their first-order theory. We also restrict attention to first-order theories F to be our foundational theories (such as ZFC). Our proposal is to analyse one dimension of the *faithfulness* or *similarity of meaning* of a formalisation as dependent upon the similarity of \mathcal{S} and what we define by R . We thus arrive at a more specific formulation of our question: what kind of similarity of \mathcal{S} and R should we aim for, or at least hope for? A notion of similarity of obvious interest in this context is elementary equivalence. Our main claim then reads as follows (see Theorem 14 for a precise statement):

Main Claim. Let F be suitable first-order foundational theory. Given a particular analysis of faithfulness in terms of elementary equivalence, an intended structure can be faithfully represented in F if and only if its (first-order) theory is decidable and F knows some decision procedure for it.

For our example structures, this implies that $(\mathbb{R}, +, \times, 0, 1, <)$ and $(\mathbb{C}, +, \times, 0, 1)$ are faithfully representable, but $(\mathbb{N}, +, \times, 0, 1, <)$ is not.

between *analytic* versus *synthetic* approaches (see, for example, [The Univalent Foundations Program, 2013], pp. 56–57, 245–246).

³In particular, we are not concerned with the relationship between informal proofs and formal derivations.

On the positive side this shows that foundational theories have an especially good grip on decidable parts of informal mathematics. Our main interest in the claim is, however, on the negative side. Many intended structures have undecidable theories, and so their study cannot be faithfully formalised in our sense. Moreover, the underlying assessment of faithfulness via elementary equivalence seems to be a fairly modest requirement on the representation of a structure, philosophically speaking.

Outline: In (§1) we recall some basics about translations between first-order theories; in particular formalising the study of some intended structure in a foundational theory. We then motivate one way of understanding the idea of a faithful such formalisation that we shall call *absolute representability*; an intended structure is absolutely representable when its representatives in models of F are elementarily equivalent to it. In (§2) we establish our main claim by proving Theorem 14. We then (§3) outline applications of our results to debates concerning first-order and higher-order resources. Finally (§4) we conclude and present some open questions.

1 Absolute representability

In this section we set up some key notions and motivate the formal definition we shall use, namely *absolute representability*.

Recall, a first-order language consists of a set of relation symbols and function symbols, each having an associated natural number called its arity; we view constant symbols as nullary function symbols. First, we fix a finite language \mathcal{L} and a first-order \mathcal{L} -structure $\mathcal{S}_{\mathcal{L}}$: this is our informal intended structure.⁴ We also fix a consistent computably enumerable first-order theory F : this is our foundational theory. We shall add another assumption on F later when needed. Examples for $\mathcal{S}_{\mathcal{L}}$ to keep in mind are $(\mathbb{N}, +, \times, 0, 1, <)$, $(\mathbb{R}, +, \times, 0, 1, <)$ or $(\mathbb{C}, +, \times, 0, 1)$, the example to keep in mind for F is ZFC; we assume ZFC is consistent.

We employ a standard definition (see e.g. [Ebbinghaus et al., 1996, Chapter VIII]) of how our intended structure $\mathcal{S}_{\mathcal{L}}$ is represented in F :⁵

⁴Here, we sidestep the discussion as to whether informal mathematics is first-order or higher-order and content ourselves with the claim that its first-order part is a substantial one. Note any restriction of attention can only strengthen the negative interpretation of our main claim.

⁵There are more general versions that allow $\psi_U(\bar{x})$ to have a tuple of free variables and/or add a formula $\psi_=_$ interpreting the equality relation; the choice is a matter of no consequence for us, our choice is for the sake of simplicity and because in many

Definition 1. A representation R of an \mathcal{L} -structure in F is a finite sequence of formulas in the language of F , namely a formula $\psi_U(x)$ such that F proves $\exists x\psi_U(x)$, and for every r -ary relation symbol $S \in \mathcal{L}$ a formula $\psi_S(x_1, \dots, x_r)$ and for every r -ary function symbol $f \in \mathcal{L}$ a formula $\psi_f(x_1, \dots, x_r, y)$ such that F proves:

$$\begin{aligned} & \forall x_1 \cdots \forall x_r \forall y \forall y' (\psi_f(x_1, \dots, x_r, y) \wedge \psi_f(x_1, \dots, x_r, y') \rightarrow y = y') \wedge \\ & \forall x_1 \cdots \forall x_r (\psi_U(x_1) \wedge \dots \wedge \psi_U(x_r) \rightarrow \exists y (\psi_U(y) \wedge \psi_f(x_1, \dots, x_r, y))) \end{aligned}$$

Such a representation definably singles out an \mathcal{L} -structure in every model of the foundational theory F as follows.

Definition 2. Let \mathcal{M} be a model of F (with universe denoted M). The \mathcal{L} -structure $R(\mathcal{M})$ has universe

$$U := \{a \in M \mid \mathcal{M} \models \psi_U(a)\}$$

and interprets an r -ary relation symbol $S \in \mathcal{L}$ by

$$\{(a_1, \dots, a_r) \in U^r \mid \mathcal{M} \models \psi_S(a_1, \dots, a_r)\},$$

and an r -ary function symbol $f \in \mathcal{L}$ by the function with the graph

$$\{((a_1, \dots, a_r), b) \in U^r \times U \mid \mathcal{M} \models \psi_f(a_1, \dots, a_r, b)\}.$$

That $R(\mathcal{M})$ is a well-defined \mathcal{L} -structure follows from the assumptions on what F proves about R in Definition 1, namely, the universe U is non-empty and ψ_f really defines the graph of some function on U .

Example 3. The usual representation $R_{\mathbb{N}}^{\text{ZFC}}$ of $(\mathbb{N}, +, \times, 0, 1, <)$ in ZFC is given taking for $\psi_U(x)$ the formula $x \in \omega$ (understood as a formula in the language $\{\in\}$ of ZFC) that defines the finite von Neumann ordinals; the formula $\psi_{<}(x_1, x_2)$ is $x_1 \in x_2$, the formulas $\psi_{+}(x_1, x_2, y)$, $\psi_{\times}(x_1, x_2, y)$ state the recursive definitions of addition and multiplication, and the formulas $\psi_0(y)$ and $\psi_1(y)$ are $y = \emptyset$ and $y = \{\emptyset\}$, respectively. The models $R_{\mathbb{N}}^{\text{ZFC}}(\mathcal{M})$, for $\mathcal{M} \models \text{ZFC}$, are called *ZFC-standard models of arithmetic* in [Hamkins and Yang, 2013]. We refer to this paper and the references therein for some information about these structures.

Given a representation R of our intended structure $S_{\mathcal{L}}$ in our foundational theory F , it is straightforward to translate first-order talk about $S_{\mathcal{L}}$ into F . The following is folklore (see e.g. [Ebbinghaus et al., 1996, Section VIII.2.E, Satz 2.2]):

foundational theories it does not matter. For example ZFC can represent all tuples by sets and quotients can be represented by sets using equivalence classes and Scott's trick.

Lemma 4. *Let R be a representation of an \mathcal{L} -structure in F . For every \mathcal{L} -sentence φ there is a sentence $R(\varphi)$ in the language of F such that for all models \mathcal{M} of F :*

$$R(\mathcal{M}) \models \varphi \iff \mathcal{M} \models R(\varphi). \quad (1)$$

Moreover, the map $\varphi \mapsto R(\varphi)$ is computable.

Proof. (Sketch) It is straightforward to compute, given a formula φ , a logically equivalent *term-reduced* formula, i.e., a formula whose atomic subformulas are of the form $x=y$, $S(\bar{x})$ or $f(\bar{x})=y$ for variables x, y, \bar{x} , relation symbols $S \in \mathcal{L}$, and function symbols $f \in \mathcal{L}$. It thus suffices to define R for term-reduced formulas. For atoms define: $R(x=y) := x=y$ and $R(S(\bar{x})) := \psi_S(\bar{x})$ and $R(f(\bar{x})=y) := \psi_f(\bar{x}, y)$. Then proceed recursively, $R(\neg\varphi) := \neg R(\varphi)$, $R(\varphi \wedge \psi) := R(\varphi) \wedge R(\psi)$ and $R(\forall x\varphi) := \forall x(\psi_U(x) \rightarrow R(\varphi))$. \square

Remark 5. The proof sketch defines $R(\neg\varphi) = \neg R(\varphi)$, a property of the map $\varphi \mapsto R(\varphi)$ that we are going to use. Slightly more generally we could use only that F proves $(R(\neg\varphi) \leftrightarrow \neg R(\varphi))$ for every φ . This follows from (1) alone: let \mathcal{M} be a model of F ; then $\mathcal{M} \models R(\neg\varphi)$ if and only if $R(\mathcal{M}) \not\models \varphi$ by (1), if and only if $\mathcal{M} \models \neg R(\varphi)$ by (1) again. Similarly, we have the equality $R(\varphi \wedge \psi) = R(\varphi) \wedge R(\psi)$ by the proof sketch and F -provable equivalence by (1) alone.

The properties of the structure $R(\mathcal{M})$ can vary significantly according to the model \mathcal{M} of F . This provides a situation in which R identifies very different structures according to the first-order model we live in. For example, what is identified by $R_{\mathbb{N}}^{\text{ZFC}}(\mathcal{M})$ can differ substantially depending on whether or not \mathcal{M} is transitive. In a transitive model $\mathcal{M} \models \text{ZFC}$, $R_{\mathbb{N}}^{\text{ZFC}}$ identifies our usual ω with the attendant operations, but if \mathcal{M} is not transitive $R_{\mathbb{N}}^{\text{ZFC}}$ may identify a structure whose domain contains non-standard natural numbers. The following question is then salient:

Question. How similar can we make our foundational representative to the intended structure? More precisely, for a given notion of similarity \sim , does there exist a representation R of \mathcal{L} -structures in F such that $\mathcal{S}_{\mathcal{L}} \sim R(\mathcal{M})$ for all models \mathcal{M} of F ?

Here, by a *notion of similarity* we mean an equivalence relation on \mathcal{L} -structures. The finer this equivalence relation, the stronger the corresponding notion of representability. Obviously, taking the identity for \sim results in an empty concept: no structure is identically representable in F . Taking isomorphism for \sim means asking whether our intended structure $\mathcal{S}_{\mathcal{L}}$ is *isomorphically representable in F* , i.e., whether

there exists a representation R such that $R(\mathcal{M}) \cong \mathcal{S}_{\mathcal{L}}$ for all $\mathcal{M} \models F$. This suggestion for \sim is naive because it is a quick consequence of the Compactness Theorem that:⁶

Proposition 6. *Only finite \mathcal{L} -structures are isomorphically representable in F .*

Hence isomorphic representability is a far too strong notion (at least as far as first-order logic is concerned). Philosophers and logicians often analyse a spectrum of similarity notions far coarser than isomorphism.⁷ We examine the prospects of choosing elementary equivalence: recall, two \mathcal{L} -structures \mathcal{A}, \mathcal{B} are *elementarily equivalent* if they satisfy the same first-order \mathcal{L} -sentences, i.e., $\text{Th}(\mathcal{A}) = \text{Th}(\mathcal{B})$, or equivalently, $\text{Th}(\mathcal{A}) \subseteq \text{Th}(\mathcal{B})$. Here, $\text{Th}(\mathcal{A})$ denotes the first-order theory of \mathcal{A} , i.e., the set of first-order sentences true in \mathcal{A} .

The corresponding notion of representability reads as follows:

Definition 7. $\mathcal{S}_{\mathcal{L}}$ is *absolutely representable in F* if there exists a representation R of an \mathcal{L} -structure in F such that $\text{Th}(\mathcal{S}_{\mathcal{L}}) = \text{Th}(R(\mathcal{M}))$ for all models \mathcal{M} of F ; in this case, we say R *absolutely represents $\mathcal{S}_{\mathcal{L}}$ in F* .

Example 8. $R_{\mathbb{N}}^{\text{ZFC}}$ from Example 3 does not absolutely represent $(\mathbb{N}, +, \times, 0, 1, <)$ in ZFC. Indeed, let $\text{Con}(\text{ZFC})$ be an arithmetical sentence expressing the consistency of ZFC. Then $\text{Con}(\text{ZFC})$ is true in $(\mathbb{N}, +, \times, 0, 1, <)$ but fails in some ZFC-standard models of arithmetic by Gödel's Second Incompleteness Theorem.

Given the naturality of elementary equivalence, the question which structures are absolutely representable in F deserves our mathematical curiosity. The above example hints at serious limitations, and we shall exactly delineate them in the next section where we establish our main claim from the Introduction. For now, we mention some reasons to find the notion philosophically interesting.

First: Clearly one goal of the informal mathematical investigation of the intended structure $\mathcal{S}_{\mathcal{L}}$ is to find out what is true in $\mathcal{S}_{\mathcal{L}}$, and first-order truth is undoubtedly an important part of it. It thus seems that an absolute representation is a clear desideratum for the foundational theory. It states that first-order truth in the intended structure does

⁶This has been noted across the structuralist literature, but is pressed particularly strongly throughout Stewart Shapiro's seminal defence of second-order logic in [Shapiro, 1991].

⁷For example, partial isomorphisms, partial isomorphism calibrated by ordinals, elementary equivalence for various logics and fragments thereof, bisimilarity, homomorphic equivalence... the list is long. What is the right similarity notion depends on the topic under consideration.

not vary with different assumptions on the model of the foundational theory we are living in.

Second: Absolute representation ensures a certain level of stability in the informal mathematical investigation of $\mathcal{S}_{\mathcal{L}}$ with respect to changes in the foundational theory. Thereby it provides comfort to the working mathematician who is not willing to restrict their investigations to F alone. For example, if we (consistently) expand F by adding more axioms, absolute representability of a structure in $\mathcal{S}_{\mathcal{L}}$ ensures that we do not change F 's beliefs about what holds in $\mathcal{S}_{\mathcal{L}}$ by doing so.⁸

Third: Absolute representability provides a reasonable way of balancing two applications of logic—what has been called the *descriptive* and *deductive* role of logic in the foundations of mathematics (see [Hintikka, 1989]). In the descriptive mode, we try to describe structures up to some level of equivalence (often isomorphism) using a logical theory. In the deductive mode, we use logic to analyse the particular kinds of *inference patterns* that appear in the relevant part of mathematics. Often these two applications of logic are argued to be in tension since there is a trade off between descriptive power and deductive completeness. Specifically, all of finiteness, natural number, real number, and various infinite well-orderings evade characterisation in first-order logic. On the other hand, logics with greater than first-order resources at their disposal are able to characterise some of these notions at the expense of pleasing meta-theoretic properties, namely compactness and Löwenheim-Skolem by Lindström's theorem [Lindström, 1969] and specifically completeness with respect to a finitary proof system. There is thus a trade-off between descriptive power and the smoothness of transition between validity and proof (we will discuss this further in §3). This tension has been formulated, for example, in [Tennant, 2000] who shows that the ability to determine models categorically is incompatible with a weak completeness requirement.⁹

One response to this predicament has been the pursuit of 'optimization projects' (cf. [Hintikka, 1989], [Read, 1997], [Dutilh Novaes, 2019])—programmes that seek to include the least amount of higher-order vocabulary in a theory necessary to ensure categoricity whilst keeping things relatively deductively well-behaved. The study of ab-

⁸It is important that we compare first-order truth of the intended structure $\mathcal{S}_{\mathcal{L}}$ and its formal counterparts $R(\mathcal{M})$ in the informal meta-language. An analogous notion *inside* $F = \text{ZFC}$ would state that $\text{Th}(R(\mathcal{M}))$ as defined in \mathcal{M} (which, assuming the universe of $R(\mathcal{M})$ is a set, can be done) does not vary with \mathcal{M} . This now includes non-standard sentences and ceases to be a property of R : this theory can vary with \mathcal{M} even when keeping $R(\mathcal{M})$ fixed; we refer to [Hamkins and Yang, 2013] for precise statements.

⁹See [Tennant, 2000], p. 267.

solute representability presents an alternative approach to optimisation along a different dimension—rather than trying small increases of higher-order resources to get categoricity whilst retaining some pleasant metatheoretic properties, we stay in first-order logic (and thereby automatically obtain these properties) and weaken the categoricity requirement to soundness and completeness regarding first-order truth.

2 Absolute representability and decidability

In this section we establish our main claim from the Introduction. We need the following lemma:

Lemma 9. *Let R be a representation of an \mathcal{L} -structure in F . Then R absolutely represents $\mathcal{S}_{\mathcal{L}}$ in F if and only if*

$$\text{Th}(\mathcal{S}_{\mathcal{L}}) = \{\varphi \mid F \vdash R(\varphi)\}. \quad (2)$$

Proof. Assume R absolutely represents $\mathcal{S}_{\mathcal{L}}$ in F . To show \subseteq in (2), let $\varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})$. We have to show that F proves $R(\varphi)$: let \mathcal{M} be a model of F , so $\varphi \in \text{Th}(R(\mathcal{M})) = \text{Th}(\mathcal{S}_{\mathcal{L}})$ by absolute representation, that is, $R(\mathcal{M}) \models \varphi$, so $\mathcal{M} \models R(\varphi)$ by Lemma 4. To show \supseteq in (2) let $\varphi \notin \text{Th}(\mathcal{S}_{\mathcal{L}})$. Then $\neg\varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})$, so F proves $R(\neg\varphi) = \neg R(\varphi)$ by the inclusion just proved. Hence $F \not\vdash R(\varphi)$ because F is consistent.

Conversely, assume (2) and let $\mathcal{M} \models F$. We have to show that $R(\mathcal{M}) \models \text{Th}(\mathcal{S}_{\mathcal{L}})$. But, by Lemma 4, $R(\mathcal{M})$ models the right-hand-side of (2). \square

Proposition 10. *If $\mathcal{S}_{\mathcal{L}}$ is absolutely representable in F , then $\text{Th}(\mathcal{S}_{\mathcal{L}})$ is decidable.*

Proof. Given as input an \mathcal{L} -sentence φ compute $R(\varphi)$ and $\neg R(\varphi) = R(\neg\varphi)$ and enumerate all consequences of F (which we assumed to be computably enumerable). By Lemma 9, exactly one of $R(\varphi)$ and $\neg R(\varphi)$ is eventually enumerated, and we accept or reject our input accordingly. \square

In Example 8 we saw that a particular representation $R_{\mathbb{N}}^{\text{ZFC}}$ is not an absolute representation. We can now say more:

Example 11. The structures $(\mathbb{N}, +, \times, 0, 1, <)$ and $(\mathbb{N}, +, \times)$, being undecidable (see [Tarski, 1968]), are not absolutely representable in F .

We now prove a partial converse to the above under an additional assumption on F : Assume there is a representation $R_{\mathbb{N}}^F$ of

$(\mathbb{N}, +, \times, 0, 1, <)$ in F such that F proves $R_{\mathbb{N}}^F(Q)$ where Q is the conjunction of the finitely many axioms of Robinson arithmetic. For any F worth calling a foundational theory, this surely is less than a minimal requirement (and clearly met by ZFC).

Given this assumption, we note the following direct consequence of Lemma 4.

Lemma 12. *Let φ be an arithmetical sentence such that Q proves φ . Then F proves $R_{\mathbb{N}}^F(\varphi)$.*

Proof. Let $\mathcal{M} \models F$. Then $\mathcal{M} \models R_{\mathbb{N}}^F(Q)$, so $R_{\mathbb{N}}^F(\mathcal{M}) \models Q$ by Lemma 4, so $R_{\mathbb{N}}^F(\mathcal{M}) \models \varphi$ as $Q \vdash \varphi$, so $\mathcal{M} \models R_{\mathbb{N}}^F(\varphi)$ by Lemma 4. \square

Let \mathbb{A} be a Turing machine. It is well-known that there is an arithmetical formula $\alpha_{\mathbb{A}}(x, y)$ with free variables x, y such that for all $n, m \in \mathbb{N}$

$$\begin{aligned} \mathbb{A}(n) = m &\implies Q \vdash \alpha_{\mathbb{A}}(\dot{n}, \dot{m}), \\ \mathbb{A}(n) \neq m &\implies Q \vdash \neg \alpha_{\mathbb{A}}(\dot{n}, \dot{m}), \end{aligned} \tag{3}$$

where \dot{n} denotes a canonical term for n (say, $\dot{0} := 0$, $\dot{1} := 1$, $\dot{2} := \dot{1} + 1$, $\dot{3} := \dot{2} + 1, \dots$).

We now present our notion of what it means for F to “know” a Turing machine deciding $\text{Th}(\mathcal{S}_{\varphi})$. Let $\ulcorner \varphi \urcorner$ denote the Gödel number of an \mathcal{L} -sentence φ .

Definition 13. Let R be a representation of an \mathcal{L} -structure in F and let \mathbb{A} be a Turing machine. F *pointwise verifies \mathbb{A} with respect to R* if for every \mathcal{L} -sentence φ :

$$F \vdash (R(\varphi) \leftrightarrow R_{\mathbb{N}}^F(\alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1}))). \tag{4}$$

We should remark here that this definition is rather weak among various reasonable notions for what it means that F “knows” some Turing machine. Note that \mathbb{A} and the inputs φ are standard (given in the meta-language), and F is asked to provide a proof of correctness “pointwise”, i.e., separately for every standard φ . Alternative notions could quantify \mathbb{A} and/or φ inside F . For example, we might require that F proves some sentence expressing “there exists a Turing machine such that for all \mathcal{L} -sentences...” where the witnessing machine might be non-standard but has to work also for nonstandard sentences.

We view Theorem 14 below as evidence that our notion of knowing a Turing machine is the right one in our context. As its proof shows, F knows in our sense *any* Turing machine deciding the theory of an absolutely representable structure, see Corollary 15.

Philosophically speaking, we might view this knowability condition as a mere technicality, and regard our result as showing that for all practical purposes absolute representability and decidability are equivalent.

The following establishes our main claim from the Introduction.

Theorem 14. *Let R be a representation of an \mathcal{L} -structure in F . Then R absolutely represents $\mathcal{S}_{\mathcal{L}}$ in F if and only if there exists a Turing machine \mathbb{A} deciding $\text{Th}(\mathcal{S}_{\mathcal{L}})$ and F pointwise verifies \mathbb{A} wrt R .*

Proof. For the forward direction, assume R absolutely represents $\mathcal{S}_{\mathcal{L}}$ in F . By Proposition 10, there is a Turing machine \mathbb{A} deciding $\text{Th}(\mathcal{S}_{\mathcal{L}})$. We claim that F pointwise verifies \mathbb{A} wrt R . Let φ be an \mathcal{L} -sentence. We show (4) by distinguishing cases.

- Case $\varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})$. Then $\mathbb{A}(\ulcorner \varphi \urcorner) = 1$ as \mathbb{A} decides $\text{Th}(\mathcal{S}_{\mathcal{L}})$, so $Q \vdash \alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1})$ by (3), so $F \vdash R_{\mathbb{N}}^F(\alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1}))$ by Lemma 12. But $F \vdash R(\varphi)$ by (2) of Lemma 9. Hence F proves both sides of the equivalence in (4).
- Case $\varphi \notin \text{Th}(\mathcal{S}_{\mathcal{L}})$. Then $\mathbb{A}(\ulcorner \varphi \urcorner) \neq 1$ as \mathbb{A} decides $\text{Th}(\mathcal{S}_{\mathcal{L}})$, so $Q \vdash \neg \alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1})$ by (3), so F proves $R_{\mathbb{N}}^F(\neg \alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1}))$ by Lemma 12. This sentence equals $\neg R_{\mathbb{N}}^F(\alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1}))$. But $\neg \varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})$, so $F \vdash \neg R(\varphi)$ using (2) and $R(\neg \varphi) = \neg R(\varphi)$. Hence F refutes both sides of the equivalence in (4).

For the converse direction, assume \mathbb{A} decides $\text{Th}(\mathcal{S}_{\mathcal{L}})$ and F pointwise verifies \mathbb{A} wrt R . We verify (2) of Lemma 9.

For \subseteq , let $\varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})$. As \mathbb{A} decides $\text{Th}(\mathcal{S}_{\mathcal{L}})$, we have $\mathbb{A}(\ulcorner \varphi \urcorner) = 1$, so $Q \vdash \alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1})$ by (3), so $F \vdash R_{\mathbb{N}}^F(\alpha_{\mathbb{A}}(\ulcorner \varphi \urcorner, \dot{1}))$ by Lemma 12, so $F \vdash R(\varphi)$ by (4).

For \supseteq , let $\varphi \notin \text{Th}(\mathcal{S}_{\mathcal{L}})$. Then $\neg \varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})$, so F proves $R(\neg \varphi) = \neg R(\varphi)$ by the above. Then $F \not\vdash R(\varphi)$ as F is consistent. \square

The proof shows:

Corollary 15. *Let R be a representation of an \mathcal{L} -structure in F and let \mathbb{A} be a Turing machine. If R absolutely represents $\mathcal{S}_{\mathcal{L}}$ in F and \mathbb{A} decides $\text{Th}(\mathcal{S}_{\mathcal{L}})$, then F pointwise verifies \mathbb{A} wrt R .*

Concerning our example structures we get:

Example 16. The structures $(\mathbb{N}, +)$, (\mathbb{N}, \times) , $(\mathbb{R}, +, \times, 0, 1, <)$ and $(\mathbb{C}, +, \times, 0, 1)$ have decidable theories [Presburger and Jabcquette, 1991, Mostowski, 1952, Tarski, 1998], and ZFC pointwise verifies their decision machines with respect to their standard representations. Hence, these structures are absolutely representable in ZFC.

3 Applications to foundational debates

We have seen thus far that an intended structure is absolutely representable in a first-order foundational theory F if and only if its (first-order) theory is decidable and F knows some decision procedure for it. In this section we'll discuss applications of this observation to the debate between proponents of first-order versus higher-order foundations.

Delineating trade-offs. An important debate in the philosophy of logic and mathematics is whether foundations should be conducted in first-order or higher-order logic (or, if one is more tolerant in outlook, which logic is suited for what purposes). Throughout this paper, we have explicitly restricted our attention to first-order theories—both with respect to the foundational theory F under consideration and the theory of informal mathematics that we are trying to formalise in F . Many authors argue that our foundational theory should contain expressive resources greater than first-order, since many notions cannot be characterised up to isomorphism in first-order logic.¹⁰ As noted in §1, all of finiteness, natural number, real number, and various infinite well-orderings evade characterisation in first-order logic whilst logics with greater than first-order resources are able to characterise more at the expense of pleasing meta-theoretic properties, especially compactness, Löwenheim-Skolem, and completeness with respect to a finitary proof system. Our results inform this trade-off by providing bounds on when a first-order foundational theory can capture truth in an intended structure. Whilst it is clearly true that for an infinite structure $\mathcal{S}_{\mathcal{L}}$, asking for isomorphic representation of $\mathcal{S}_{\mathcal{L}}$ is too much, nonetheless our results show that there are precise conditions on which a first-order theory can be omniscient concerning truth in $\mathcal{S}_{\mathcal{L}}$. This shows that for a certain class of infinite structures, even first-order logic can have a good deal of traction (structurally speaking), in contrast to the view that only finite structures can be well-treated in first-order logic. Moreover, we obtain this traction precisely when the theory is decidable and F knows some decision procedure for it. We now discuss some payoffs of these observations for debates concerning the roles of first-order and higher-order logics.

Technical correlates of epistemic arguments. Our main result provides a technical correlate of a kind of epistemic argument one can

¹⁰This is argued in various places, but [Shapiro, 1991] is one of the strongest advocates of the use of higher-order resources for certain purposes in foundations.

find in the literature. [Read, 1997] puts forward the following epistemic objection to the use of higher-order logic (albeit in order to reject it later):

Without deductive completeness, knowledge of any mathematical treatment—indeed of any theory—would be impossible. Only through completeness can claims be tested and verified. If someone is challenged concerning a claim made as part of a complete theory, production of a proof can settle the matter – where ‘proof’ is understood as in first-order logic, as a decidable property, so that any putative proof can be checked for correctness. If the theory was not complete, one might claim that a thesis of it was true but not provable. How could such a claim be checked and supported or refuted? ([Read, 1997], p. 88)

Of course, as Read notes, the objection is not sound since it appears to motivate decidability as the operative property rather than completeness:

Even in pure first-order logic, completeness guarantees only semi-decidability—that we can find a proof if there is one. It does not yield a corresponding effective method of refutation. The argument for the epistemological need for completeness in fact suggests that decidability is really the desired property; but that feature, though dreamed of, perhaps, by Leibniz and Wittgenstein, was denied us even for first-order consequence in the aftermath of Gödel’s incompleteness theorem. ([Read, 1997], p. 88)

Our results show that, from the point of view of an appropriate F there is a technical manifestation of this argument put forward for consideration in [Read, 1997]. As it turns out (from the point of view of F) we have traction on truth in a structure $\mathcal{S}_{\mathcal{L}}$ exactly when F knows a Turing machine witnessing the decidability of the theory of $\mathcal{S}_{\mathcal{L}}$. Thus, if one (controversially) wishes to maintain that the decidability criterion is the important one for epistemic tractability, one obtains a good match between truth in a structure and satisfaction of this epistemic criterion for those structures with decidable theories, and poor epistemic traction on truth when the theory of $\mathcal{S}_{\mathcal{L}}$ is undecidable.

Limitations of first-order foundations. On the other side of the coin, we have shown that a foundation which is both first-order and computably enumerable has limits in absolutely representing theories.

The meta-theoretic advantages given by compactness, Löwenheim-Skolem, and completeness have their price. Not only will any first-order foundational theory F fail to determine the cardinality of an intended infinite structure $\mathcal{S}_{\mathcal{L}}$, but if the theory of $\mathcal{S}_{\mathcal{L}}$ is undecidable F loses traction on truth in $\mathcal{S}_{\mathcal{L}}$ too. Authors such as [Tennant, 2000] show that weak completeness requirements on a theory are incompatible with the categorial representability of a structure in that theory. We have shown further that if we want traction on truth in intended structures with undecidable theories, we will *need* higher-order resources.¹¹ This in turn shows that if we want to use some foundational theory to provide a *generous arena* and *shared standard* (in [Maddy, 2019]’s sense) accuracy concerning truth for structures with undecidable theories requires non-first-order resources.

Implications for optimization projects. The previous observations have implications for so-called ‘optimization projects’. In §1 we noted that various authors (e.g. [Hintikka, 1989], [Read, 1997], [Dutilh Novaes, 2019]) consider the idea of trying to find the ‘mildest’ strengthenings of first-order logic in order to be able to capture certain structures up to isomorphism whilst retaining many desirable meta-theoretic properties. We suggested absolute representability as a different approach to optimisation, instead of increasing the strength of the logic to obtain categoricity, we weaken the level of accuracy concerning structural description required and coarsen the similarity relation to elementary equivalence. Our results indicate that this optimisation project can be successful for precisely those structures which have decidable theories, but a different approach is needed for structures with undecidable ones.

4 Conclusion

Recall, we asked for a representation R such that $\mathcal{S}_{\mathcal{L}}$ and $R(\mathcal{M})$ are similar for all models \mathcal{M} of F . Taking similarity as elementary equivalence, we saw that this requires that $\text{Th}(\mathcal{S}_{\mathcal{L}})$ be decidable. We have analysed some implications for studying intended structures via absolute representability. If $\text{Th}(\mathcal{S}_{\mathcal{L}})$ is not decidable, however, it is natural to ask for weaker notions of representability.

There are many possibilities and we briefly discuss one of them, namely the one obtained by weakening the equality in (2) of Lemma 9

¹¹This observation coheres well with suggestions of [Shapiro, 1991] and [Väänänen, 2001].

to an inclusion: call R a *sound representation of $\mathcal{S}_{\mathcal{L}}$ in F* if

$$\{\varphi \mid F \vdash R(\varphi)\} \subseteq \text{Th}(\mathcal{S}_{\mathcal{L}}). \quad (5)$$

Roughly said, F proves only true first-order sentences about $\mathcal{S}_{\mathcal{L}}$. Clearly, the working mathematician studying $\mathcal{S}_{\mathcal{L}}$ would reject any foundational theory not providing such a representation.

Proposition 17. *Let R be a representation of \mathcal{L} -structures in F . Then R is a sound representation of $\mathcal{S}_{\mathcal{L}}$ in F if and only if there exists a model \mathcal{M} of F such that $R(\mathcal{M})$ is elementarily equivalent to $\mathcal{S}_{\mathcal{L}}$.*

Proof. Assume (5) holds. It suffices to show that the theory

$$F \cup \{R(\varphi) \mid \varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})\}$$

is consistent. Indeed, a model \mathcal{M} of this theory has the property that $R(\mathcal{M}) \models \varphi$ for all $\varphi \in \text{Th}(\mathcal{S}_{\mathcal{L}})$ by Lemma 4, so $R(\mathcal{M})$ and $\mathcal{S}_{\mathcal{L}}$ are elementarily equivalent. The claimed consistency follows from compactness: if the theory above is inconsistent, then there are finitely many $\varphi_1, \dots, \varphi_k \in \text{Th}(\mathcal{S}_{\mathcal{L}})$ such that F refutes $R(\varphi_1) \wedge \dots \wedge R(\varphi_k) = R(\psi)$ for $\psi := \varphi_1 \wedge \dots \wedge \varphi_k$ (see Remark 5); as F proves $\neg R(\psi) = R(\neg\psi)$ and $\neg\psi \notin \text{Th}(\mathcal{S}_{\mathcal{L}})$, this contradicts (5).

Conversely, if (5) fails, then there exists $\varphi \notin \text{Th}(\mathcal{S}_{\mathcal{L}})$ such that F proves $R(\varphi)$. By Lemma 4, $R(\mathcal{M}) \models \varphi$ for every model \mathcal{M} of F while $\mathcal{S}_{\mathcal{L}} \not\models \varphi$, so $R(\mathcal{M})$ and $\mathcal{S}_{\mathcal{L}}$ are not elementarily equivalent. \square

Thus, asking for a sound representation is asking for a special model \mathcal{M} of F , namely one such that $R(\mathcal{M})$ and $\mathcal{S}_{\mathcal{L}}$ are elementarily equivalent. This makes sense also for other notions of similarity, and in particular for isomorphism. For example, in the case of $(\mathbb{N}, +, \times, 0, 1, <)$ the latter asks F to have an ω -model. It is thus philosophically justified to ask F to be more than just consistent, namely to ask for the existence of some model \mathcal{M} of F such that $R(\mathcal{M})$ and $\mathcal{S}_{\mathcal{L}}$ are similar in some or another sense for various intended structures $\mathcal{S}_{\mathcal{L}}$. But it is unclear how much one should or can ask for.

Another way to weaken the notion of representability, in order to make it apply to structures with undecidable theories, is to consider only “intended” models \mathcal{M} of F , or an expansion thereof formulated using greater than first-order resources.¹² For example, restricting \mathcal{M} to transitive standard models of $F = \text{ZFC}$ makes $R_{\mathbb{N}}^{\text{ZFC}}$ an isomorphic

¹²A strategy along these lines is considered by [Väänänen, 2001] who suggests that we work with *urlogic*—our underlying logic that, in a manner somewhat reminiscent of Frege, carries its intended meaning with it.

representation of $(\mathbb{N}, +, \times, 0, 1, <)$; on the other hand, $(\mathbb{R}, +, \times, 0, 1, <)$ is absolutely but *not* isomorphically representable. By contrast, if we instead formulate ZFC in quasi-weak second-order logic (where second-order variables are stipulated to range over *countable* relations), $(\mathbb{R}, +, \times, 0, 1, <)$ becomes *isomorphically* representable. We suggest that various “semantic” extensions of F can, in principle, be philosophically assessed by their capability to represent intended structures with respect to one or another notion of similarity.¹³ This project, while worthwhile, is very broad, and so we leave the study of such similarity relations in semantic extensions as an open problem.

References

- [Barton, F] Barton, N. (F). Structural relativity and informal rigour. In Boscolo, S., Ternullo, C., and Oliveri, G., editors, *Objects, Structures, and Logics: FilMat Studies in the Philosophy of Mathematics*, Boston Studies in the Philosophy and History of Science. Springer. Forthcoming.
- [Dutilh Novaes, 2019] Dutilh Novaes, C. (2019). Axiomatizations of arithmetic and the first-order/second-order divide. *Synthese*, 196(7):2583–2597.
- [Ebbinghaus et al., 1996] Ebbinghaus, H.-D., Flum, J., and Thomas, W. (1996). *Einführung in die Mathematische Logik*. Spektrum Akademischer Verlag, Heidelberg, Berlin. 4. Auflage.
- [Hamkins and Yang, 2013] Hamkins, J. D. and Yang, R. (2013). Satisfaction is not absolute. arXiv:1312.0670v1 [math.LO].
- [Hintikka, 1989] Hintikka, J. (1989). Is there completeness in mathematics after Gödel? *Philosophical Topics*, 17(2):69–90.
- [Kennedy et al., S] Kennedy, J., Magidor, M., and Väänänen, J. (S). Inner models from extended logics: Part 1. Manuscript under review.
- [Lindström, 1969] Lindström, P. (1969). On extensions of elementary logic. *Theoria*, 35(1):1–11.
- [Maddy, 2019] Maddy, P. (2019). What do we want a foundation to do? In Centrone, S., Kant, D., and Sarikaya, D., editors, *Reflections on the Foundations of Mathematics: Univalent Foundations, Set Theory and*

¹³Some semantic extensions of ZFC, and what is isomorphically representable therein, are considered in [Shapiro, 2001], [Barton, F], and [Kennedy et al., S].

- General Thoughts*, pages 293–311. Springer International Publishing, Cham.
- [Mostowski, 1952] Mostowski, A. (1952). On direct products of theories. *J. Symbolic Logic*, 17(1):1–31.
- [Presburger and Jabcquette, 1991] Presburger, M. and Jabcquette, D. (1991). On the completeness of a certain system of arithmetic of whole numbers in which addition occurs as the only operation. *History and Philosophy of Logic*, 12(2):225–233.
- [Rav, 2008] Rav, Y. (2008). The axiomatic method in theory and in practice. *Logique et Analyse*, 51(202):125–147.
- [Read, 1997] Read, S. (1997). Completeness and categoricity: Frege, Gödel and model theory. *History and Philosophy of Logic*, 18(2):79–93.
- [Shapiro, 1991] Shapiro, S. (1991). *Foundations without Foundationalism: A Case for Second-order Logic*. Oxford University Press.
- [Shapiro, 2001] Shapiro, S. (2001). Systems between first-order and second-order logics. In Gabbay, D. M. and Guenther, F., editors, *Handbook of Philosophical Logic*, pages 131–187. Springer Netherlands, Dordrecht.
- [Tarski, 1968] Tarski, A. (1968). *Undecidable Theories*. Amsterdam: North-Holland Pub. Co.
- [Tarski, 1998] Tarski, A. (1998). A decision method for elementary algebra and geometry. In Caviness, B. F. and Johnson, J. R., editors, *Quantifier Elimination and Cylindrical Algebraic Decomposition*, pages 24–84, Vienna. Springer Vienna.
- [Tennant, 2000] Tennant, N. (2000). Deductive versus expressive power: A pre-Gödelian predicament. *The Journal of Philosophy*, 97(5):257–277.
- [The Univalent Foundations Program, 2013] The Univalent Foundations Program (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. <http://homotopytypetheory.org/book>.
- [Väänänen, 2001] Väänänen, J. (2001). Second-order logic and foundations of mathematics. *The Bulletin of Symbolic Logic*, 7(4):504–520.