Tracing Internal Categoricity[*]

Jouko Väänänen[†]

Department of Mathematics and Statistics,
University of Helsinki

Institute for Logic, Language and Computation,
University of Amsterdam

March 7, 2020

Abstract

Informally speaking, the categoricity of an axiom system means that its non-logical symbols have only one possible interpretation that renders the axioms true. Although non-categoricity has become ubiquitous in the second half of the 20th century whether one looks at number theory, geometry or analysis, the first axiomatizations of such mathematical theories by Dedekind, Hilbert, Huntington, Peano and Veblen were indeed categorical. A common resolution of the difference between the earlier categorical axiomatizations and the more modern non-categorical axiomatizations is that the latter derive their non-categoricity from Skolem's Paradox and Gödel's Incompleteness Theorems, while the former, being second order, suffer from a heavy reliance on metatheory, where the Skolem-Gödel phenomenon re-emerges. Using second order meta-theory to avoid non-categoricity of the meta-theory would only seem to lead to an infinite regress. In this paper we maintain that internal categoricity breaks this traditional picture. It applies to both first and second order axiomatizations, although in the first order case we have so far only examples. It does not depend on the meta-theory in a way that would lead to an infinite regress. And it covers the classical categoricity results of early researchers. In the first order case it is weaker than categoricity itself, and in the second order case stronger. We give arguments suggesting that internal categoricity is the "right" concept of categoricity.

1 Introduction

The concept of categoricity of a formal theory was introduced with this name by Veblen [18], who notes that the name was suggested to him by John Dewey. Veblen introduced

1

this concept in the context of geometry, but it turned out to have a vastly more general applicability. Veblen defines:

> Inasmuch as the terms *point* and *order* are undefined one has a right, in thinking of the propositions, to apply the terms in connection with any class of objects of which the axioms are valid propositions. It is part of our purpose however to show that there is *essentially only one* class of which the twelve axioms are valid. In more exact language, any two classes K and K' of objects that satisfy the twelve axioms are capable of a one-to-one correspondence such that if any three elements A, B, C of K are in the order ABC, the corresponding elements of K' are also in the order ABC. Consequently any proposition which can be made in terms of points and order either is in contradiction with our axioms or is equally true of all classes that verify our axioms. . . . A system of axioms such as we have described is called *categorical* . . .

In modern terminology we would say that the twelve axioms of Veblen are categorical in the following sense:

Definition 1. *A set of axioms*[1] *is categorical if any two of its models are isomorphic.*

Veblen points out what we would now call the *completeness* of categorical axiom systems: for any given sentence ϕ in the vocabulary of the axioms, either ϕ follows[2] from the axioms or $\neg\phi$ follows. As Veblen notes, the concept had been already used with a different name by Huntington [7], who calls it 'sufficiency', and by Hilbert [5], who is less explicit about it. For an excellent survey of the history of categoricity and completeness, we refer to [1].

If L is a vocabulary, i.e. a set of relation, function and constant symbols, an L-model (or an L-structure) $\mathfrak{M}$ consists of a domain M, which is a non-empty set, together with an interpretation of the symbols of L in the set M. Respectively an L-sentence of first (or second) order logic is a sentence the non-logical symbols of which are in L. The interpretation of a relation symbol R is denoted $R^{\mathfrak{M}}$, the interpretation of a function symbol f is denoted $f^{\mathfrak{M}}$, and the interpretation of a constant symbol c is denoted $c^{\mathfrak{M}}$. Isomorphism $\mathfrak{M} \cong \mathfrak{M}'$ of L-structures $\mathfrak{M}$ and $\mathfrak{M}'$ is defined in the usual way. In this paper identity $=$ is considered a logical symbol and therefore we do not explicitly mention whether identity is included in the vocabulary or not. The interpretation of identity is always equality. If $\mathfrak{M}$ is an L-structure and $L' \subseteq L$, then $\mathfrak{M} \upharpoonright L'$ denotes the reduct of $\mathfrak{M}$ to the vocabulary L'.

The most famous as well as historically the first categorical axiomatization of a mathematical structure is Dedekind's second order axiomatization of elementary arithmetic. Note that in the second order setup the arithmetic operations of addition and multiplications are definable from the successor relation.

---

[1] For this definition to make sense it does not matter in which formal language the axioms are written as long as the concepts of *model* and *isomorphism* make sense. In this paper the axioms are first or second order axioms.

[2] In modern terminology Veblen's concept of "follows" would be perhaps best called a semantic consequence relation.

Example 2 (Dedekind)**.** *Let us consider the vocabulary $\{s, 0\}$ of the theory of the successor function on the natural numbers. The second order sentence $N^2(s, 0)$:*

$$\begin{aligned}
&\forall xy(s(x) = s(y) \to x = y) \wedge \\
&\forall x \neg s(x) = 0 \wedge \\
&\forall X((X(0) \wedge \forall x(X(x) \to X(s(x)))) \to \forall x X(x))
\end{aligned} \tag{1}$$

is categorical with the unique (up to isomorphism) model $(\mathbb{N}, f, 0)$, where $f(x) = x + 1$.

Since in second order logic we can existentially quantify over the function s and the constant 0, we can axiomatize the countably infinite domain:

Example 3. *The second order sentence $I^2 = \exists F \exists x N^2(F, x)$ is categorical with the unique (up to isomorphism) model*[3] $\mathbb{N}$.

Example 2 is different from Example 3 in the following important respect. If we take two models of $N^2(s, 0)$, there is a *unique* isomorphism between them. But if we take two models of the sentence I^2 of Example 3, there is an isomorphism (i.e. a bijection) between the models but the isomorphism is by no means unique, because the set $\mathbb{N}$ permits a continuum of different bijections.

Example 4. *We consider the vocabulary $\{R, \varepsilon\}$, where R is unary and ε is binary. The second order sentence P^2:*

$$\begin{aligned}
&I^2 \text{ relativized (see below for the definition of relativization) to } R \wedge \\
&\forall xy(x\varepsilon y \to R(x)) \wedge \\
&\forall xy(\forall z(z\varepsilon x \leftrightarrow z\varepsilon y) \to x = y) \wedge \\
&\forall X \exists x \forall y(R(y) \to (X(y) \leftrightarrow y\varepsilon x))
\end{aligned} \tag{2}$$

is categorical with the unique (up to isomorphism) model $(\mathcal{P}(\mathbb{N}), \in, \mathbb{N})$, where $\in$ is the usual membership-relation between elements of $\mathbb{N}$ and elements of $\mathcal{P}(\mathbb{N})$.

Again we may note that the isomorphisms manifesting the categoricity of P^2 are not unique because of the many automorphisms of the R-part. However, if the isomorphism is fixed on the R-parts, the rest is unique.

Example 5 (Huntington)**.** *The second order axiomatization R^2 of the completely ordered field of real numbers, i.e. the conjunction of the first order axioms of ordered fields and the following* Least Upper Bound Principle*:*

$$\begin{aligned}
&\forall X((\exists x X(x) \wedge \exists y \forall x(X(x) \to x < y)) \to \\
&\quad \exists y(\forall x(X(x) \to (x < y \vee x = y)) \wedge \\
&\quad\quad \forall y'(\forall x(X(x) \to (x < y' \vee x = y')) \to (y < y' \vee y = y'))))\text{,}
\end{aligned} \tag{3}$$

is categorical with the unique (up to isomorphism) model $(\mathbb{R}, +, \cdot, 0, 1, <)$. The usual proof of the categoricity proceeds by first isolating the natural numbers as $0, 1, 1 + 1, 1 + 1 + 1, \ldots$, then the rationals, and then the reals as the completion of the rationals. The isomorphism between any two models of R^2 is unique.

---

[3]Of the empty vocabulary.

As a final example, let us consider set theory:

Example 6 (Zermelo). *Let ZFC^2 be the conjunction of the second order Zermelo-Fraenkel axioms, obtained from the ordinary Zermelo-Fraenkel axioms by replacing the Separation and Replacement Schemas by their second order versions. Models of ZFC^2 are, up to isomorphism, of the form $(V_\kappa, \in)$, where $\kappa > \omega$ is inaccessible. Since there may be many inaccessibles, ZFC^2 is not categorical unless we make large cardinal assumptions. However, as emphasised already by Zermelo [20], ZFC^2 is categorical in the weaker sense that if the height of the model is fixed, then there is, up to isomorphism, only one model. This weak form of categoricity of ZFC^2 is sometimes called* quasi-categoricity.

The range of categoricity among second order theories is extensive. In fact, it is very hard to find a structure which would be not second order characterizable, without using a cardinality argument or the Axiom of Choice, see [8].

2 Preliminaries

We considered above some examples of categoricity. Let us now set the stage for a more general approach. The remarkable property of second order logic is that it can express the categoricity of its own sentences. To see what this means we have to introduce some notation.

Let us consider a finite vocabulary $L = \{R_1, \ldots, R_\alpha, f_1, \ldots, f_\beta\}$ and a unary predicate U not in L. If $\mathfrak{M}$ is an $L \cup \{U\}$-structure, let the L-structure $\mathfrak{M}^U$ be the relativization of $\mathfrak{M} \restriction L$ to the interpretation of the predicate U. For $\mathfrak{M}^U$ to be a legitimate L-structure, something has to be assumed about the interpretation of U as well as about the interpretations of the function symbols in $\mathfrak{M}$. Let $\text{Res}_{\{f_1, \ldots, f_\beta\}}(U)$ be the conjunction of the first order sentences

$$\begin{aligned}
&\exists x_1 U(x_1)\\
&\forall x_1 \ldots x_n((U(x_1) \wedge \ldots \wedge U(x_n)) \to U(f_j(x_1, \ldots, x_n))),
\end{aligned} \tag{4}$$

where $j \in \{1, \ldots, \beta\}$. Certainly, $\mathfrak{M}$ satisfies $\text{Res}_{\{f_1, \ldots, f_\beta\}}(U)$ if and only if $\mathfrak{M}^U$ is an L-structure. If ϕ is a second order sentence, we let ϕ^U denote the result of relativizing ϕ to U. Thus in ϕ^U all first order quantifiers are restricted to range over elements of U, and the second order variables are restricted to range over subsets of U, relations over U and functions on U. In consequence, if ϕ is an L-sentence and $\mathfrak{M} \models \text{Res}_{\{f_1, \ldots, f_\beta\}}(U)$, then

$$\mathfrak{M} \models \phi^U \iff \mathfrak{M}^U \models \phi. \tag{5}$$

Let $L' = \{R'_1, \ldots, R'_\alpha, f'_1, \ldots, f'_\beta\}$ be another vocabulary, where the arity of each R'_i is the same as the arity of R_i and the same for the function symbols. We assume $L \cap L' = \emptyset$. If ϕ is a second order sentence, we use ϕ' to denote the result of replacing each R_i by R'_i and f_j by f'_j in ϕ. Let U' be a new unary predicate symbol not in $L \cup L' \cup \{U\}$. Let $\text{ISO}_{L,L'}(F, U, U')$ be the second order sentence which says that the

function F defines an isomorphism between the L-part relativized to U and the L'-part relativized to U', i.e. $\mathrm{ISO}_{L,L'}(F, U, U')$ is the conjunction of the first order sentences

$$\forall x(U(x) \rightarrow U'(F(x)))$$
$$\forall xy(F(x) = F(y) \rightarrow x = y)$$
$$\forall x(U'(x) \rightarrow \exists y(U(y) \wedge F(y) = x))$$
$$\forall x_1 \ldots x_n(R_i(x_1, \ldots, x_n) \leftrightarrow R'_i(F(x_1), \ldots, F(x_n)))$$
$$\forall x_1 \ldots x_n(F(f_j(x_1, \ldots, x_n)) = f'_j(F(x_1), \ldots, F(x_n))),$$

where $i \in \{1, \ldots, \alpha\}$ and $j \in \{1, \ldots, \beta\}$.

Definition 7. *For any second order L-sentence ϕ, the second order[4] sentence CAT_ϕ of the empty vocabulary is defined as follows:*

$$CAT_\phi : \ \forall UU'R_1 \ldots R_\alpha R'_1 \ldots R'_\alpha f_1 \ldots f_\beta f'_1 \ldots f'_\beta CAT_\phi^+.$$

where CAT_ϕ^+ is

$$(Res_{\{f_1, \ldots, f_\beta\}}(U) \wedge Res_{\{f'_1, \ldots, f'_\beta\}}(U') \wedge \phi^U \wedge \phi'^{U'}) \rightarrow \exists F\, \mathrm{ISO}_{L,L'}(F, U, U').$$

Lemma 8. *Suppose ϕ is a second order L-sentence. The following conditions are equivalent:*

(C1) *ϕ is categorical.*

(C2) *CAT_ϕ is valid in the empty vocabulary.*

(C3) *CAT_ϕ^+ is valid in the vocabulary $L \cup L' \cup \{U, U'\}$.*

Proof. To see this, suppose first (C1). We show that CAT_ϕ is valid. Suppose M is an arbitrary non-empty set. We can consider M the domain of a model of the empty vocabulary i.e. the vocabulary of CAT_ϕ. Let $\mathfrak{M}$ be the $L \cup L' \cup \{U, U'\}$-structure resulting from interpreting the non-logical symbols of $L \cup L' \cup \{U, U'\}$ in the domain M in some arbitrary way. Let us assume $\mathfrak{M}$ satisfies $\mathrm{Res}_{\{f_1, \ldots, f_\beta\}}(U) \wedge \mathrm{Res}_{\{f'_1, \ldots, f'_\beta\}}(U') \wedge \phi^U \wedge \phi'^{U'}$. Thus, by (5), $(\mathfrak{M} \restriction L \cup \{U\})^U$ is an L-structure satisfying ϕ. Respectively, $(\mathfrak{M} \restriction L' \cup \{U'\})^{U'}$ is an L'-structure satisfying ϕ'. Changing vocabularies, the categoricity of ϕ yields $\mathfrak{M} \models \exists F\, \mathrm{ISO}_{L,L'}(F, U, U')$. Clearly, (C2) implies (C3). Finally, assume (C3). Let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ be two L-models of ϕ. By changing the vocabulary, we can translate $\mathfrak{M}_2$ into an L'-model $\mathfrak{M}'_2$ of ϕ'. Since $L \cap L' = \emptyset$, we can form an $L \cup L' \cup \{U, U'\}$-structure $\mathfrak{M}$ such that $(\mathfrak{M} \restriction L \cup \{U\})^U = \mathfrak{M}_1$ and $(\mathfrak{M} \restriction L' \cup \{U'\})^{U'} = \mathfrak{M}'_2$. By (C3), $\mathfrak{M} \models \exists F\, \mathrm{ISO}_{L,L'}(F, U, U')$, whence $\mathfrak{M}_1 \cong \mathfrak{M}_2$. $\square$

It should be noted that even though CAT_ϕ has no non-logical symbols, there is a marked difference between CAT_ϕ being true in a model (of the empty vocabulary) and it being valid in all models (of the empty vocabulary). There is nothing surprising in

---

[4]For simplicity, we use same names for second order relation (or function) variables and for (first order) relation and function symbols.

this. Truth in a model is, in logic in general, very far from being equivalent to truth in all models. However, the case of CAT_ϕ is a very special one.

The truth of CAT_ϕ in a model of (the empty vocabulary) of cardinality κ means the same as categoricity of ϕ in models of cardinality $\leq \kappa$. In other words, despite its appearance, the sentence CAT_ϕ does not state the categoricity of ϕ, because of size limitations. It is only the proposition *CAT_ϕ is valid* which states it.

For the categoricity of ϕ it obviously suffices that CAT_ϕ has arbitrarily large models. If we let κ denote the Hanf-number[5] of second order logic then it suffices that CAT_ϕ is true in a model of cardinality κ of the empty vocabulary. Since the Hanf-number of second order logic is less than the first extendible cardinal ([10]), for ϕ to be categorical it suffices that CAT_ϕ has a model of size at least the first extendible cardinal.

3   Internal categoricity

The idea behind internal categoricity is the observation that in familiar cases of categorical second order sentences ϕ the sentence CAT_ϕ is not only valid but even provable. This is remarkable because second order logic does not have a similar Completeness Theorem as first order logic. So there is no a priori reason why a valid sentence would be provable.

Following [6] we include in the axioms of second order logic the following schema: Suppose L is a vocabulary. The *L-Axiom Schema of Comprehension* is the following set of second order sentences:

$$\forall x_1 \ldots x_n \forall X_1 \ldots X_k \exists X \forall y_1 \ldots y_m (X(y_1, \ldots, y_m) \leftrightarrow \\ \phi(x_1, \ldots, x_n, X_1, \ldots, X_k, y_1, \ldots, y_m)), \tag{6}$$

where $\phi(x_1, \ldots, x_n, X_1, \ldots, X_k, y_1, \ldots, y_m)$ is an arbitrary second order L-formula not containing X free. By provability in second order logic we mean provability from the axioms of second order logic including (6). The schema (6) is non-trivial even when $\phi(x_1, \ldots, x_n, X_1, \ldots, X_k, y_1, \ldots, y_m)$ is first order.

Definition 9. *Suppose L is a vocabulary. A second order L-sentence ϕ is* internally categorical *if the sentence CAT_ϕ (or equivalently, CAT_ϕ^+) is provable in second order logic.*

Since provable sentences are valid, internal categoricity implies categoricity. Moreover, internal categoricity can be established for important categorical sentences:

Theorem 10. *The following second order sentences are internally categorical:*

1. *The axiomatization N^2 of $(\mathbb{N}, s, 0)$. ([11], [3], [13], [19]).*

2. *The axiomatization I^2 of $\mathbb{N}$ in the empty vocabulary.*

3. *The axiomatization P^2 of $(\mathcal{P}(\mathbb{N}), \in, \mathbb{N})$.*

---

[5]The Hanf number of a logic is the least cardinal such that if a sentence of the logic has a model of size at least κ, it has arbitrarily large models.

4. *The axiomatization R^2 of $(\mathbb{R}, +, \cdot, 0, 1, <)$.*

Proof. For a soft proof one can use Theorem 12 below. In all cases the proof is the classical proof of categoricity. When one examines the classical proof carefully, nothing beyond the axioms of second order logic is used. Of course, (6) is heavily used, although not in its full generality. □

The idea of the internal categoricity of N^2 appears already in [11]. In fact, when categoricity of a sentence ϕ is defined in [9] (English translation [14]), it is not defined as the truth of CAT_ϕ, but as the provability of CAT_ϕ in the simple theory of types. We go further and require provability in second order logic itself. The phrase 'internal categoricity' was introduced in the context of arithmetic in [19]. It was advocated more generally in [15], with more details in [17]. We refer to [2] for a recent discussion on this concept.

The point of internal categoricity is that it establishes categoricity without requiring a set-theoretic background (meta-)theory. Internal categoricity of, say N^2 or R^2, is an "internal" property of second order logic itself. Likewise, the proof of CAT_ϕ gives finite evidence of the categoricity of ϕ. Again, there is no need to refer to set theory or "full semantics" (see below). The appealing contention that the categoricity of second order axiomatizations of familiar mathematical structures requires "full" second order logic, with set-theoretical semantics, would be simply wrong.

Although second order logic does not have a Completeness Theorem in the same sense as first order logic, there is a more general concept of semantics, due to Henkin [4], which permits a Completeness Theorem. In this more general semantics we do not let bound second order variables, such as $U, U', R_i, R'_i, g_i, f'_i, F$ in CAT_ϕ above, range over *all* possible relations (and functions) over the given domain M but just over elements of a set $\mathcal{G}$ of relations (and functions) on M. This is in line with the development in first order logic, made completely explicit in the proof of Gödel's Completeness Theorem, to consider arbitrary models, with arbitrary relations, functions and elements to interpret the non-logical symbols, even if one is interested in number theory, where the functions symbols $+$ and $\times$ have a clear intended meaning. In second order logic the intended meaning of the range of second order variables (for subsets of the domain) is the entire power-set of the domain, but we allow the more general set $\mathcal{G}$ as the range in order to achieve a Completeness Theorem.

Definition 11 ([4]). *Suppose L is a vocabulary. The combination $(\mathfrak{M}, \mathcal{G})$ of an L-structure $\mathfrak{M}$ and a set $\mathcal{G}$ satisfying the L-Axiom Schema of Comprehension is called a* Henkin L-model.

Intuitively speaking the set $\mathcal{G}$ of a Henkin L-model $(\mathfrak{M}, \mathcal{G})$ is closed in the sense that any set or relation that can be defined from the interpretation of the symbols of L by means of quantification over $\mathcal{G}$ is actually already in $\mathcal{G}$. The schema (6) is quite strong even in the case that $L = \emptyset$.

It goes without saying that the set of *all* subsets and relations of a given domain M satisfies (6). Likewise, it is obvious that in any Henkin model $(\mathfrak{M}, \mathcal{G})$ the set $\mathcal{G}$ is a Boolean algebra which contains all finite subsets of M, as well as the interpretations of

the symbols of L. Because of the strong impredicativity[6] of (6) it is a non-trivial task to construct Henkin models satisfying (6). The method of Henkin [4] yields examples of such $(\mathfrak{M}, \mathcal{G})$ with countable $\mathcal{G}$ and M.

The point of Henkin models is:

Theorem 12 ([4])**.** *Suppose L is a vocabulary. A second order L-sentence is provable if and only if it is valid in all Henkin L-models.*

By combining Definition 9 and Theorem 12 we get:

Theorem 13. *Suppose L is a vocabulary. A second order L-sentence ϕ is* internally categorical *if the sentence CAT_ϕ is true in all Henkin models.*

Theorem 13 explains why internal categoricity is called "internal": It is *internal* in the sense of being internal to each $\mathcal{G}$, namely, it is required that if the interpretations of the symbols of the vocabulary are in $\mathcal{G}$, then so is the interpretation of F. Nothing is claimed about interpretations outside $\mathcal{G}$. Since Henkin semantics satisfies the Compactness Theorem, the sentences N^2 and R^2 are not categorical *across* Henkin models, i.e. there are Henkin models of e.g. N^2 which are non-standard. For every fixed $\mathcal{G}$ there is a unique model of N^2, but different $\mathcal{G}$ may give rise to non-isomorphic models.

For second order set theory ZFC^2 internal categoricity holds only in the weaker sense of quasi-categoricity explained in Example 6. However, for example, for

$$ZFC^2 + \text{"there are no inaccessible cardinals} > \omega\text{"}$$

internal categoricity can be proved [17].

Our list of internally categorical second order sentences is only the beginning. There are many more examples. In fact, it would be rather surprising if there were an example of a mathematical structure, constructed without the use of Axiom of Choice, a cardinality argument, or the enumeration techniques à la Gödel, which did not have an internally categorical second order characterization.

In sum, the classical categoricity results of second order logic, due to Dedekind, Veblen and others, hold in the stronger sense of internal categoricity. The advantage of internal categoricity is that it has always a finite proof in the formal language of second order logic itself. Therefore it does not depend on set-theoretical meta-theory in the same sense as ordinary categoricity.

4 First order internal categoricity—mapping the landscape

Because of the extraordinary ability of second order logic to express the categoricity of its own sentences, it would seem that internal categoricity, based on the assumption of the categoricity statement holding in all Henkin models, has no role in first order logic.

---

[6]The relation X that is defined by $\phi(x_1, \ldots, x_n, X_1, \ldots, X_k, y_1, \ldots, y_m)$ occurs in the range of the universally and existentially quantified second order variables in $\phi(x_1, \ldots, x_n, X_1, \ldots, X_k, y_1, \ldots, y_m)$.

However, this is far from true. We will show that first order arithmetic and first order set theory have *a form* of internal categoricity, reminiscent of the internal categoricity of their second order cousins. But we do not (know how to) define a general concept of first order internal categoricity. Rather, we are mapping the landscape in search of a good definition. At the moment we have just a few examples.

First order theories are not categorical *per se*. Their defining characteristic is the existence of models of all infinite cardinalities and in many important cases also the existence of non-standard models in each infinite cardinality separately. First order internal categoricity has to circumvent these undeniable facts.

Let us start with first order Peano arithmetic. Let $P(+, \cdot)$ be the infinite set of Peano axioms with the Schema of Induction

$$
\begin{aligned}
&\forall x_1 \ldots x_n((\phi(0, x_1, \ldots, x_n) \wedge \\
&\forall y(\phi(y, x_1, \ldots, x_n) \rightarrow \phi(y+1, x_1, \ldots, x_n))) \rightarrow \\
&\forall y \phi(y, x_1, \ldots, x_n)),
\end{aligned} \tag{7}
$$

where 0 and 1 are defined terms denoting the identity elements of $+$ and $\cdot$, respectively. Ordinarily we assume that the formula $\phi(y, x_1, \ldots, x_n)$ of (7) is any first order formula of the vocabulary $\{+, \cdot\}$ of arithmetic. However, given a vocabulary L, let us use $P(+, \cdot, L)$ to denote the extension of $P(+, \cdot)$ where the Induction Schema (7) is formulated for first order formulas $\phi(y, x_1, \ldots, x_n)$ of the vocabulary $\{+, \cdot\} \cup L$.

Theorem 14 (Internal categoricity of Peano arithmetic). *Suppose $\mathfrak{M}$ is a model of*

$$
P(+, \cdot, \{+', \cdot'\}) \cup P(+', \cdot', \{+, \cdot\}). \tag{8}
$$

Then there is $\pi : \mathfrak{M} \restriction \{+, \cdot\} \cong \mathfrak{M} \restriction \{+', \cdot'\}$. Moreover, the mapping π is first order definable on $\mathfrak{M}$.

Proof. Note that $\mathfrak{M}$ may very well be a non-standard model of any infinite cardinality. Let 0 and 1 be the first elements of $\mathfrak{M} \restriction \{+, \cdot\}$, and respectively $0', 1'$. Let $\psi(x, u, v)$ say that x codes, using $+$ and $\cdot$, an initial segment I with the last element u, of $\mathfrak{M} \restriction \{+, \cdot\}$, an initial segment I' with the last element v, of $\mathfrak{M} \restriction \{+', \cdot'\}$, and a function $f : I \rightarrow I'$ such that $f(0) = 0'$, $f(y+1) = f(y) +' 1'$ for all $y \in I \setminus \{u\}$, and $f(u) = v$. Let $\phi(u, v)$ be the formula $\exists x \psi(x, u, v)$.

First we show by induction on a that for every $a \in M$ there is $b \in M$ such that $\mathfrak{M} \models \phi(a, b)$. If $a = 0$, then we need only code the pair $(0, 0')$ to make $\phi(0, 0')$ true. Suppose then $\phi(a, b)$ is true. Let x, I, I' and f be as in $\psi(x, a, b)$. Clearly, $f(1) = 1'$. Let I_1 be I with $a + 1$ added, and similarly I'_1. Extend f to f_1 by mapping the last element $a + 1$ of I_1 to the last element $b +' 1'$ of I'_1. By coding I_1, I'_1 and f_1 we get x_1 such that $\psi(x_1, a + 1, b +' 1')$. Thus $\phi(a + 1, b +' 1')$ is true. By the Induction Schema, since $\mathfrak{M}$ satisfies $P(+, \cdot, \{+', \cdot'\})$, for every $a \in M$ there is b such that $\mathfrak{M} \models \phi(a, b)$. Analogously, since $\mathfrak{M}$ satisfies $P(+', \cdot', \{+, \cdot\})$, we can prove by induction in $\mathfrak{M} \restriction \{+', \cdot'\}$ that for every $b \in M$ there is a such that $\mathfrak{M} \models \phi(a, b)$.

Next we show by induction on a that if $\phi(a, b)$ and $\mathfrak{M} \models \phi(a, b')$, then $b = b'$. Suppose first $\phi(0, b)$ and $\phi(0, b')$. Let x, I, I' and f be as in $\psi(x, 0, b)$, and x_1, I_1, I'_1 and f' as in $\psi(x_1, 0, b')$. Then $b = f(0) = 0'$ and $b' = f'(0) = 0'$. Thus $b = b'$. Suppose then $\phi(a + 1, b) \wedge \phi(a + 1, b')$ is true. Let x, I, I' and f be as in $\psi(x, a + 1, b)$,

and x_1, I_1, I_1' and f' as in $\psi(x_1, a+1, b')$. Since we assume that a satisfies the claim, $f(a) = f'(a)$. Hence $b = f(a+1) = f(a) +' 1' = f'(a) +' 1' = f'(a+1) = b'$. By the Induction Schema, since $\mathfrak{M}$ satisfies $P(+, \cdot, \{+', \cdot'\})$, for every $a \in M$ there is a unique b such that $\mathfrak{M} \models \phi(a, b)$.

Note that if $\phi(a, a')$ and $\phi(b, b')$ with x, I, I' and f witnessing $\psi(x_1, a, a')$ and x_1, I_1, I_1' and f' witnessing $\psi(x_1, b, b')$, then $f(y) = f'(y)$ for $y \in I \cap I'$. This is because the claim clearly holds for $y = 0$, and if it holds for y, then $f(y + 1) = f(y) +' 1' = f'(y) +' 1' = f'(y + 1)$. So the claim follows by induction in $\mathfrak{M} \restriction \{+, \cdot\}$ as $\mathfrak{M}$ satisfies $P(+, \cdot, \{+', \cdot'\})$.

Finally, we show that the relation $F = \{(a, b) \in M \times M : \mathfrak{M} \models \phi(a, b)\}$ is an isomorphism $\mathfrak{M} \restriction \{+, \cdot\} \to \mathfrak{M} \restriction \{+', \cdot'\}$. Suppose α, β and γ are such that $\alpha + \beta = \gamma$. Suppose x, I, I' and f witness $\psi(x, \gamma, f(\gamma))$. By the above, $f \subseteq F$. We use induction on β in $\mathfrak{M} \restriction \{+, \cdot\}$, to prove that $f(\alpha) +' f(\beta) = f(\gamma)$, appealing to the fact that $\mathfrak{M}$ satisfies $P(+, \cdot, \{+', \cdot'\})$. If $\beta = 0$, then $\alpha = \gamma$ and

$$f(\alpha) +' f(\beta) = f(\alpha) +' f(0) = f(\alpha) +' 0' = f(\alpha) = f(\gamma).$$

Assume then $\beta = \delta + 1$. Now

$$f(\alpha) +' f(\beta) = f(\alpha) +' f(\delta + 1) = f(\alpha) +' f(\delta) +' 1' =$$

$$= f(\alpha + \delta) +' 1' = f(\alpha + \delta + 1) = f(\alpha + \beta) = f(\gamma).$$

Suppose then α, β and γ are such that $\alpha \cdot \beta = \gamma$. We use induction on β in $\mathfrak{M} \restriction \{+, \cdot\}$ to prove that $f(\alpha) \cdot' f(\beta) = f(\gamma)$, appealing to the fact that $\mathfrak{M}$ satisfies $P(+, \cdot, \{+', \cdot'\})$. If $\beta = 0$, then $\gamma = 0$ and

$$f(\alpha) \cdot' f(\beta) = f(\alpha) \cdot' f(0) = f(\alpha) \cdot' 0' = 0' = f(\gamma).$$

Assume then $\beta = \delta + 1$. Now

$$f(\alpha) \cdot' f(\beta) = f(\alpha) \cdot' f(\delta + 1) = f(\alpha) \cdot' (f(\delta) +' 1') =$$

$$= f(\alpha) \cdot' f(\delta) +' f(\alpha) \cdot' 1' = f(\alpha \cdot \delta) +' f(\alpha) = f(\alpha \cdot \delta + \alpha) = f(\gamma).$$

$\square$

Since π above is definable and "$\pi : \mathfrak{M} \restriction \{+, \cdot\} \cong \mathfrak{M} \restriction \{+', \cdot'\}$" is a first order statement—let us denote it $\mathrm{ISOM}(+, \cdot, +', \cdot')$—we can rephrase the above theorem as a theorem of first order logic:

Corollary 15 (Internal categoricity of Peano arithmetic in first order logic)**.**

$$P(+, \cdot, \{+', \cdot'\}) \cup P(+', \cdot', \{+, \cdot\}) \vdash \mathrm{ISOM}(+, \cdot, +', \cdot'). \tag{9}$$

The internal categoricity of first order arithmetic may seem simply false on the basis that there are a continuum of non-isomorphic countable models of Peano's axioms. Indeed, suppose $(\mathbb{N}, +', \cdot')$ is a non-standard model of Peano arithmetic such that $+'$ and $\cdot'$ are Δ_2^0-definable (by [12]) as relations. Then $(\mathbb{N}, +, \cdot, +', \cdot')$ satisfies $P(+, \cdot, \{+', \cdot'\})$ and $P(+', \cdot')$, but there is no reason to think that it satisfies

$P(+', \cdot', \{+, \cdot\})$, for when the Δ_2^0-definable relations $x +' y = z$ and $x \cdot' y = z$ are constructed, only induction for formulas involving $+'$ and $\cdot'$ are (or can be) considered. We know that $(\mathbb{N}, +', \cdot')$ cannot satisfy N^2, so it is interesting to note that the failure of the Induction Axiom of (1) in $(\mathbb{N}, +', \cdot')$ is manifested by the failure to satisfy $P(+', \cdot', \{+, \cdot\})$.

We now move to set theory. The first order analogue of ZFC^2 is the Zermelo-Fraenkel axiom system ZFC. Let us denote the vocabulary of this theory $\{\varepsilon_1\}$. To formulate internal categoricity of ZFC we introduce another binary symbol ε_2 and compare ZFC in the vocabulary $\{\varepsilon_1\}$, denoted $ZFC(\varepsilon_1)$, and ZFC in the vocabulary $\{\varepsilon_2\}$, denoted $ZFC(\varepsilon_2)$. If we allow symbols from a new vocabulary L to occur in the Separation Schema and the Replacement Schema of $ZFC(\varepsilon_1)$, we denote the extended theory $ZFC(\varepsilon_1, L)$. Similarly $ZFC(\varepsilon_2, L)$.

The second order theory ZFC^2 is not categorical per se. One has to fix the (inaccessible) cardinality of the model. In the first order context we accomplish this by assuming that the two models have the same domain.

Theorem 16 (Internal categoricity of first order ZFC, [16])**.** *Suppose $\mathfrak{M}$ is a model of*

$$ZFC(\varepsilon_1, \{\varepsilon_2\}) \cup ZFC(\varepsilon_2, \{\varepsilon_1\}).$$

Then there is $\pi : \mathfrak{M} \upharpoonright \{\varepsilon_1\} \cong \mathfrak{M} \upharpoonright \{\varepsilon_2\}$. Moreover, the mapping π is first order definable on $\mathfrak{M}$.

We may again ask, what about the incompleteness of ZFC? If we start with a countable transitive model (M, ε_1) of ZFC, we may use inner models or forcing to get other countable model (M', ε_2) of ZFC with the same order-type of the class of ordinals. By applying a re-enumeration we may assume $M' = M$. Numerous set-theoretic statements, such as the Continuum Hypothesis CH or the Souslin Hypothesis SH, may be true (or false) in (M, ε_1) but false (or true) in (M, ε_2). How is this utter incompleteness consistent with the idea of internal categoricity? The answer to this riddle is the following: An important part of the proof that an inner model is a model of ZFC, or that a forcing extension preserves all the axioms of ZFC, is checking the validity of the Separation Schema and the Replacement Schema. Just as in the case of the Henkin construction for a non-standard model, we simply cannot allow extra non-logical symbols in the formulas of the Separation Schema and the Replacement Schema. If, on the other hand, we aim at the second order version ZFC^2 with bound relation variables instead of formulas in the Separation the Replacement Axioms, we cannot get the independence results.

Both arithmetic and set theory manifest internal categoricity via a definable isomorphism. This means that the statement of internal categoricity is first order expressible just as categoricity of a second order sentence is expressible in second order logic itself.

5   What is the value of internal categoricity?

Categoricity is such a beautiful and simple concept that one hesitates to modify it in any way. However, categoricity is usually defined in set theory and may also *depend*

on set theory. We have argued that internal categoricity is a more basic concept that does *not* depend on set theory.

In the second order case internal categoricity is stronger than categoricity and can be rephrased simply as provability of the statement of categoricity, because categoricity can be expressed in the second order language itself. In the first order case the situation is more complex. For one thing, we cannot directly express categoricity by first order means. Conceivably we could still consider the provability of the set-theoretic statement of categoricity but there are no non-trivial examples among first order theories. Instead, the internal categoricity approach is to require categoricity only as far as an alternative model can be 'seen'. In the case of first order internal categoricity we offer an alternative model, such as $+', \cdot', \epsilon'$, to be 'seen' by adding it to the language and allowing it to be used in the axioms. If the axioms remain true in the extended language the power of the axioms generates an isomorphism.

Loosely speaking, first order arithmetic and set theory are categorical in their own vicinity, among models they can see, but they are not categorical if models are allowed to be constructed from 'outside' at will. The difference to second order arithmetic and set theory is that the bound relation variables of the latter reach all possible models, in the case of Henkin models all models coded by the Henkin model. In a sense, measured by the Henkin model, they see all possible models, all possible models are in their vicinity, and still their axioms are true. That explains why they are categorical. Still the ultimate reason for their categoricity is that they possess the property of being internally categorical. In this sense internal categoricity is the "right" concept of categoricity, "right" in the sense of being the most fundamental.

References

[1] Steve Awodey and Erich H. Reck. Completeness and categoricity. part I: Nineteenth-century axiomatics to twentieth-century metalogic. *History and Philosophy of Logic*, 23(1):1–30, 2002.

[2] Tim Button and Sean Walsh. Structure and categoricity: Determinacy of reference and truth value in the philosophy of mathematics. *Philosophia Mathematica*, 24(3):283–307, 2016.

[3] Solomon Feferman and Geoffrey Hellman. Predicative foundations of arithmetic. *Journal of Philosophical Logic*, 24(1):1–17, 1995.

[4] Leon Henkin. Completeness in the theory of types. *J. Symbolic Logic*, 15:81–91, 1950.

[5] David Hilbert. Grundlagen der Geometrie. Zweite, durch Zusätze vermehrte und mit fünf Anhängen versehene Auflage. Leipzig: B. G. Teubner. V u. 175 S. gr. 8° (1903)., 1903.

[6] David Hilbert and Wilhelm Ackermann. *Grundzüge der theoretischen Logik*. Springer-Verlag, Berlin, 1972. Sixth edition, Die Grundlehren der mathematischen Wissenschaften, Band 27. First edition published 1928.

[7] E. V. Huntington. A complete set of postulates for the theory of of absolute continuous magnitude. *Trans. Am. Math. Soc.*, 3:264–279, 1902.

[8] Tapani Hyttinen, Kaisa Kangas, and Jouko Väänänen. On second-order characterizability. *Log. J. IGPL*, 21(5):767–787, 2013.

[9] Adolf Lindenbaum and Alfred Tarski. Über die Beschränktheit der Ausdrucksmittel deduktiver Theorien. Erg. Math. Kolloqu. 7, 15-22 (1936)., 1936.

[10] Menachem Magidor. On the role of supercompact and extendible cardinals in logic. *Israel J. Math.*, 10:147–157, 1971.

[11] Charles Parsons. The structuralist view of mathematical objects. *Synthese*, 84(3):303–346, 1990.

[12] Hilary Putnam. Trial and error predicates and the solution to a problem of Mostowski. *J. Symbolic Logic*, 30:49–57, 1965.

[13] Stephen G. Simpson and Keita Yokoyama. Reverse mathematics and Peano categoricity. *Ann. Pure Appl. Logic*, 164(3):284–293, 2013.

[14] Alfred Tarski. *Logic, semantics, metamathematics*. Hackett Publishing Co., Indianapolis, IN, second edition, 1983. Papers from 1923 to 1938, Translated by J. H. Woodger, Edited and with an introduction by John Corcoran.

[15] Jouko Väänänen. Second order logic or set theory? *Bulletin of Symbolic Logic*, 18(1):91–121, 2012.

[16] Jouko Väänänen. An extension of a theorem of Zermelo. *Bulletin of Symbolic Logic*, 25, 2019.

[17] Jouko Väänänen and Tong Wang. Internal categoricity in arithmetic and set theory. *Notre Dame J. Form. Log.*, 56(1):121–134, 2015.

[18] Oswald Veblen. A system of axioms for geometry. *Trans. Am. Math. Soc.*, 5:343–384, 1904.

[19] James Walmsley. Categoricity and indefinite extensibility. *Proceedings of the Aristotelian Society*, 102(3):217–235, 2002.

[20] Ernst Zermelo. Über Grenzzahlen und Mengenbereiche. Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundam. Math.*, 16:29–47, 1930.