An introduction to internal categoricity

Jouko Väänänen

Helsinki, May 2022

European Research Council
Established by the European Commission
ACADEMY OF FINLAND

Recent paper: *Philosophical Uses of Categoricity Arguments*, with Penelope Maddy, 62 pages. To appear. Available in arXiv.

Natural numbers

In 1893 Dedekind gave an axiomatisation/definition of the structure of natural numbers:

$$\mathbb{N} = \{0, 1, 2, \ldots\}, s(n) = n + 1$$

Dedekind's axiomatization

- $s(n) \neq 0$
- $s(n) = s(m) \rightarrow n = m$
- Induction Principle: If A is a set of natural numbers such that A contains 0 and is closed under the function s, then $A = \mathbb{N}$.

Typical second order categoricity results

Theorem (Dedekind 1893)

If $(M_1, s_1, 0_1)$ and $(M_2, s_2, 0_2)$ satisfy Dedekind's second order axioms, then $(M_1, s_1, 0_1) \cong (M_2, s_2, 0_2)$.

Theorem (Zermelo 1930)

If $(M_1, \in_1)$ and $(M_2, \in_2)$ satisfy the second order Zermelo-Fraenkel axioms[1] and $|M_1| = |M_2|$, then $(M_1, \in_1) \cong (M_2, \in_2)$.

[1] Without urelements, for simplicity.

Is this set theory?

- Second order categoricity results seem to depend on a (non-absolute) set–theoretical background semantics[2].
- Can mathematics be built on second order logic alone?

[2] due to the interpretation of the second order quantifiers.

Internal categoricity

- In the classical cases the axioms of second order logic suffice for categoricity.
- No non-absolute set-theoretical semantics is needed.
- I call this “internal categoricity”, a term first used[3] by Walmsley (2002).

[3] in number theory.

Some building blocks

- Comprehension Schema: $\exists P \forall \vec{x}(P(\vec{x}) \leftrightarrow \varphi(\vec{x}, \vec{y}, \vec{Q}))$, where P is not free in $\varphi(\vec{x}, \vec{y}, \vec{Q})$.
- Categoricity is written in second order logic itself, following Lindenbaum &Tarski 1936.
- Provability can be expressed in terms of Henkin-semantics, which is absolute[4].

[4] unlike validity.

A detailed definition

- T_1 a second-order theory, voc. $\{R_1, \ldots, R_n\}$, R_i r_i-ary.
- $\{R'_1, \ldots, R'_n\}$ new relation symbols, R'_i r_i-ary.
- T_2 is T_1 with every occurrence of R_i replaced by R'_i, $1 \leq i \leq n$.
- F a new unary function symbol, U and U' new unary predicate symbols.
- $\mathrm{ISO}(F, U, U')$ says (in first order logic) that F is a bijection between U and U' such that $R_i(a_1, \ldots, a_{r_i}) \leftrightarrow R'_i(F(a_1), \ldots, F(a_{r_i}))$ for all $a_1, \ldots, a_n \in U$ and all i with $1 \leq i \leq n$.

Recall: relativization

- If φ is any second-order formula, $\varphi^{(U)}$ is obtained from φ by relativizing all first- and second-order quantifiers to U.
- For any theory T, $T^{(U)}$ denotes $\{\varphi^{(U)} : \varphi \in T\}$.

Definition

The second order theory T_1 is *internally categorical* if

$$T_1^{(U)} \cup T_2^{(U')} \vdash_2 \exists F\, \mathrm{ISO}(F, U, U').$$

- The (usual) categoricity of T_1 is equivalent to

$$T_1^{(U)} \cup T_2^{(U)} \models \exists F\, \mathrm{ISO}(F, U, U').$$

- Internal categoricity says that the categoricity of T_1 is not merely a set-theoretical fact but is, in fact, provable in second-order logic.

Naturally, internal categoricity is stronger than (the usual) categoricity:

Theorem

Suppose T_1 is internally categorical. Then T_1 is categorical.

The (trivial) proof

- W.l.o.g., suppose the domains of M and M', which we denote also with M and M', are disjoint.
- Let M^* be the unique model with $M \cup M'$ as domain, $U^{M^*} = M$, $U'^{M^*} = M'$, $R_i^{M^*} = R_i^M$, and $R'^{M^*}_i = R_i^{M'}$ for $1 \leq i \leq n$. Clearly, $M^* \models T_1^{(U)} \cup T_2^{(U')}$.
- Hence by internal categoricity and the Soundness Theorem, $M^* \models \exists F\, \mathrm{ISO}(F, U, U')$.
- Hence, $(U^{M^*}, R_1^{M^*}, \ldots, R_n^{M^*}) \cong (U'^{M^*}, R'^{M^*}_1, \ldots, R'^{M^*}_n)$.
- Now $M \cong M'$ follows.

- Semantically, internal categoricity says that every Henkin model of $T_1^{(U)} \cup T_2^{(U')}$ satisfies $\exists F\, \mathrm{ISO}(F, U, U')$.
- If a Henkin model recognizes two models of T_1, it also recognizes an isomorphism between them.
- Note: $\mathrm{ISO}(F, U, U')$ is absolute in Henkin models.
- In every Henkin model there is at most one (up to $\cong$) model of T_1.

- If T_1 has an infinite model, then (by the Compactness Theorem) there are Henkin models $H = (M, \mathcal{A})$ of T_1 of any infinite cardinality.
- The Henkin models of T_1 form a "cloud" $\mathcal{H}(T_1)$ around the full model M of (an internally categorical) T_1.
- Inside each "point" H of the cloud $\mathcal{H}(T_1)$ there is a unique model M_H of T_1 in the sense of H.
- Truth of a sentence φ in $M \models T_1$ has a stronger (than $M \models \varphi$) and more absolute version, viz. truth in each $M_H \in \mathcal{H}(T_1)$, i.e. $T_1 \vdash_2 \varphi$.

Internal categoricity in number theory

When T_1 is PA^2, we have

Theorem (e.g. [Hel89][5])
PA^2 is internally categorical.

Hence

Corollary (Dedekind)
PA^2 is categorical.

[5] [Par90], [Sha91], [Wal02], probably others.

Internal categoricity of PA^2—semantic version

Corollary

If $(M_1, s_1, 0_1)$ and $(M_2, s_2, 0_2)$ satisfy Dedekind's second order axioms in a Henkin model H, then $(M_1, s_1, 0_1) \cong (M_2, s_2, 0_2)$, and the isomorphism is in H.

In other words: In every Henkin model there is at most one (up to $\cong$) model of PA^2.

Proof

By the Π^1_1-Comprehension Schema, there is

$$R = \bigcap \{P : P(0_1, 0_2) \wedge \forall x \in M_1 \forall y \in M_2 (P(x,y) \rightarrow P(s_1(x), s_2(y)))\}.$$

It is easy to prove that R is the desired isomorphism.

Much weaker principles suffice: Simpson and Yokoyama (2013) show that WKL_0 (Weak König's Lemma[6]) suffices over RCA_0 (Recursive Comprehension Axiom). This is interesting because WKL_0 is weaker than Peano arithmetic itself (i.e. ACA_0).

[6] Every infinite binary subtree of $2^{<\omega}$ has an infinite branch.

Internal categoricity of PA^2—syntactic version

Let $\mathbf{PA}^2_1$ be the conjunction of the axioms of PA^2 in the vocabulary $\{0_1, S_1\}$. Similarly $\mathbf{PA}^2_1$ in $\{0_2, S_2\}$.

Theorem (Internal categoricity restated)

$\vdash_2 \forall N_1 0_1 S_1 N_2 0_2 S_2((\mathbf{PA}_1^{2(N_1)} \wedge \mathbf{PA}_2^{2(N_2)}) \to \exists F\, \mathrm{ISO}(F, N_1, N_2))$.

Corollary

Suppose $\varphi(0, S)$ *is a second-order sentence in the vocabulary* $\{0, S\}$*. Then*

$$\vdash_2 \forall N_1 0_1 S_1 N_2 0_2 S_2((\mathbf{PA}_1^{2(N_1)} \wedge \mathbf{PA}_2^{2(N_2)})$$

$$\rightarrow (\varphi^{(N_1)}(0_1, S_1) \leftrightarrow \varphi^{(N_2)}(0_2, S_2))).$$

A detailed proof of this can be found in [BW18].

Theorem ([BW18] 'Intolerance' of $\mathbf{PA}^2$)

Suppose $\varphi(0, S)$ is a second-order sentence. Then

$$\vdash_2 \forall N 0 S(\mathbf{PA}^{2(N)} \to \varphi^{(N)}(0, S)) \vee \forall N 0 S(\mathbf{PA}^{2(N)} \to \neg\varphi^{(N)}(0, S))).$$

In other words, the class of all Henkin models is divided into those in which the unique model (if any) of PA^2 satisfies φ and to those in which it satisfies $\neg\varphi$.

How is this different from $PA^2 \vdash \varphi \vee \neg\varphi$? The class of all possible models of PA^2 being divided into those which satisfy φ and those which satisfy $\neg\varphi$.

Proof.
Corollary 9 says

$$\vdash \forall N_1 0_1 S_1 N_2 0_2 S_2 ((\mathbf{PA}_1^{2(N_1)} \wedge \mathbf{PA}_2^{2(N_2)})$$

$$\rightarrow (\varphi^{(N_1)}(0_1, S_1) \vee \neg\varphi^{(N_2)}(0_2, S_2))).$$

By rearranging quantifiers and connectives, we obtain

$$\vdash \forall N_1 0_1 S_1 (\mathbf{PA}_1^{2(N_1)} \rightarrow \varphi^{(N_1)}(0_1, S_1))$$

$$\vee \forall N_2 0_2 S_2 (\mathbf{PA}_2^{2(N_2)} \rightarrow \neg\varphi^{(N_2)}(0_2, S_2)))$$

from which the claim follows by change of bound variables. □

Corollary ([BW18])

If φ is a second-order sentence in the vocabulary $\{0, S\}$, then

$$\mathbf{PA}_1^{2(N_1)} \cup \mathbf{PA}_2^{2(N_2)} \cup \{\varphi^{(N_1)}(0_1, S_1), \neg\varphi^{(N_2)}(0_2, S_2)\}$$

is deductively inconsistent.

How is this different from the inconsistency of $\mathrm{PA}^2 \cup \{\varphi\} \cup \{\neg\varphi\}$?

What does internal categoricity of PA^2 give us?

- It is stronger than (or as strong as) categoricity.
- Its proof is more absolute than the proof of categoricity.
- It is (?) the actual content of the categoricity of PA^2.

Internal categoricity in set theory

Internal categoricity in set theory

Theorem ([VW15])

If $(M_1, \in_1)$ and $(M_2, \in_2)$ satisfy the second order Zermelo-Fraenkel axioms[7] and $|M_1| = |M_2|$ in a Henkin model H, then $(M_1, \in_1) \cong (M_2, \in_2)$, and the isomorphism is in H.

(Proof below)

[7] Urelements are irrelevant for our purposes, so for simplicity, we assume there are none.

- ZF^2, the usual vocabulary $\{\in\}$.[8]
- E_1 and E_2 binary relation symbols and X_1 and X_2 unary relation symbols.
- If φ is a second-order sentence in the vocabulary $\{\in\}$, let $\varphi(E_1)$ and $\varphi(E_2)$ be translations of φ into the vocabularies $\{E_1\}$ and $\{E_2\}$, respectively.

[8] Urelements are irrelevant for our purposes, so for simplicity, we assume there are none.

- $\mathbf{ZF}^{2(X_1)}(E_1) = \bigwedge\{\varphi^{(X_1)}(E_1) : \varphi(E_1) \in \mathrm{ZF}^2(E_1)\}$
  $\mathbf{ZF}^{2(X_2)}(E_2) = \bigwedge\{\varphi^{(X_2)}(E_2) : \varphi(E_2) \in \mathrm{ZF}^2(E_2)\}$.
- Let IA be the second-order sentence in the vocabulary $\{X_1, E_1, X_2, E_2\}$ which says that the classes of inaccessible cardinals in the sense of (X_1, E_1) and (X_2, E_2), respectively, are isomorphic.
- Let $\mathrm{ISO}(F, X_1, X_2)$ be the first-order sentence that says F is an isomorphism (X_1, E_1) between (X_2, E_2).

Internal quasi-categoricity of ZF^2

Theorem ([VW15])
$\vdash_2 (\mathbf{ZF}^{2(X_1)}(E_1) \wedge \mathbf{ZF}^{2(X_2)}(E_2) \wedge IA) \rightarrow \exists F\,\mathrm{ISO}(F, X_1, X_2).$

Every Henkin model has at most one (up to $\cong$) model of set theory with prescribed inaccessibles.

This implies Zermelo's quasi-categoricity theorem[9].

[McG97] and [Lav99] present a “schematic" version of this.

[9] in the absence of urelements.

Proof

Let π be an isomorphism between the inacessibles of (X_1, E_1) and (X_2, E_2). Let

$$\begin{aligned} R &= \bigcap \{P : \pi \subseteq P \wedge \forall x \in X_1 \forall y \in X_2 \\ & ((\forall z E_1 x \exists u E_2 y P(z,u) \wedge \forall u E_2 y \exists z E_1 x P(z,u)) \\ & \rightarrow P(x,y))\}. \end{aligned}$$

It is easy to prove that R is the desired isomorphism.

Corollary

Suppose φ *is a second-order sentence in* $\{\in\}$. *Then*

$$\vdash (\mathbf{ZF}^{2(X_1)}(E_1) \wedge \mathbf{ZF}^{2(X_2)}(E_2) \wedge IA) \rightarrow (\varphi^{(X_1)}(E_1) \leftrightarrow \varphi^{(X_2)}(E_2)).$$

A detailed proof of this can be found in [BW18].

- Let IA_0 be the first-order sentence of set theory saying that there are no inaccessible cardinals, i.e. that every limit cardinal $> \omega$ is singular.
- Let
$$\Gamma = \bigwedge(\mathrm{ZF}^2 \cup \{\mathrm{IA}_0\}).$$

Theorem ([BW18])

('Intolerance') Suppose φ *is a second-order sentence in* $\{\in\}$*. Then*

$$\vdash_2 \forall X \forall E(\Gamma^{(X)}(E) \to \varphi^{(X)}(E)) \vee \forall X \forall E(\Gamma^{(X)}(E) \to \neg\varphi^{(X)}(E)).$$

How is this different from $ZF^2 \vdash \varphi \vee \neg\varphi$? The class of all Henkin models of ZF^2 is divided into those which satisfy φ and those which satisfy $\neg\varphi$.

Proof.

By Corollary 14,

$$\vdash_2 \forall X_1 \forall E_1 \forall X_2 \forall E_2((\Gamma^{(X_1)}(E_1) \wedge \Gamma^{(X_2)}(E_2))$$

$$\rightarrow (\varphi^{(X_1)}(E_1) \vee \neg\varphi^{(X_2)}(E_2))).$$

By rearranging quantifiers and connectives, we obtain

$$\vdash_2 \forall X_1 \forall E_1(\Gamma^{(X_1)}(E_1) \rightarrow \varphi^{(X_1)}(E_1))$$

$$\vee \forall X_2 \forall E_2(\Gamma^{(X_2)}(E_2) \rightarrow \neg\varphi^{(X_2)}(E_2)))$$

from which the claim follows by change of bound variables. □

Corollary ([BW18])

If φ is a second-order sentence in vocabulary $\{\in\}$, then the theory

$$\{\Gamma^{(X_1)}(E_1), \Gamma^{(X_2)}(E_2), \varphi^{(X_1)}(E_1), \neg\varphi^{(X_2)}(E_2)\}$$

is deductively inconsistent.

How is this different from the fact that there is no Henkin model of $ZF^2 \cup \{\varphi\} \cup \{\neg\varphi\}$?

What does internal quasi-categoricity of ZF^2 give us?

- It is stronger than (or as strong as) quasi-categoricity.
- Its proof is more absolute than the proof of quasi-categoricity.
- It is (?) the actual content of the quasi-categoricity of ZF^2.

Internal categoricity of first order theories: arithmetic

A framework for two domains

- Let N_1 and N_2 be unary predicate symbols, $+_1$ and $+_2$, $\cdot_1$ and $\cdot_2$ binary function symbols, and 0_1 and 0_2, 1_1 and 1_2 constant symbols.
- Let $\mathrm{PA}_1(N_1)$ be the first-order theory PA written in the vocabulary $\{+_1, \cdot_1, 0_1, 1_1\}$, with the functions $+_1, \cdot_1$ mapping $N_1 \times N_1$ to N_1, the constants $0_1, 1_1$ in N_1, the Induction Schema allowing formulas from the larger vocabulary $\{+_1, \cdot_1, 0_1, 1_1\} \cup \{+_2, \cdot_2, 0_2, 1_2\}$ and (first-order) quantifiers ranging over the whole domain, including $N_1 \cup N_2$.
- Likewise $\mathrm{PA}_2(N_2)$.

Arithmetic on two different domains.

Remark

$PA_1(N_1) \cup PA_2(N_2)$ does not imply the isomorphism of the N_1-part and the N_2-part.

- Let M be the disjoint sum of two copies of the standard model $(\mathbb{N}, 0, S)$. Thus $(N_1^M, 0_1^M, S_1^M) \cong (\mathbb{N}, 0, S)$ and $(N_2^M, 0_2^M, S_2^M) \cong (\mathbb{N}, 0, S)$.
- Let $(\mathbb{N}^*, 0^*, S^*)$ be a countable non-standard model elementarily equivalent to (but again disjoint from) $(\mathbb{N}, 0, S)$ and let M' be the disjoint sum of $(\mathbb{N}, 0, S)$ and $(\mathbb{N}^*, 0^*, S^*)$, making it a model of the vocabulary $\mathrm{PA}_1(N_1) \cup \mathrm{PA}_2(N_2)$.
- A simple Ehrenfeucht-Fraïssé-game argument, as in Feferman [Fef72], shows that $M \equiv M'$.
- Since $(\mathbb{N}, 0, S)$ satisfies even the second-order Induction Axiom, M certainly satisfies $\mathrm{PA}_1(N_1)$. Respectively, M satisfies $\mathrm{PA}_2(N_2)$. Thus $M \models \mathrm{PA}_1(N_1) \cup \mathrm{PA}_2(N_2)$ and then also $M' \models \mathrm{PA}_1(N_1) \cup \mathrm{PA}_2(N_2)$. But $(N_1^{M'}, 0_1^{M'}, S_1^{M'})$ and $(N_2^{M'}, 0_2^{M'}, S_2^{M'})$ are non-isomorphic, as the former is standard and the latter is non-standard.

A new framework for two domains

- The vocabulary of PA is $\{+, \cdot, 0, 1\}$.
- Let N_1 and N_2 be unary predicate symbols, $+_1$ and $+_2$, $\cdot_1$ and $\cdot_2$ binary function symbols, and 0_1 and 0_2, 1_1 and 1_2 constant symbols.
- Let $PA_1(N_1)$ be the first-order theory PA written in the vocabulary $\{+_1, \cdot_1, 0_1, 1_1\}$, with the functions $+_1, \cdot_1$ mapping $N_1 \times N_1$ to N_1, the constants $0_1, 1_1$ in N_1, the Induction Schema allowing formulas from the larger vocabulary $\{+, \cdot, 0, 1\} \cup \{+_1, \cdot_1, 0_1, 1_1\} \cup \{+_2, \cdot_2, 0_2, 1_2\}$ and (first-order) quantifiers ranging over the whole domain, including $N_1 \cup N_2$.
- Likewise $PA_2(N_2)$.

- An obvious obstacle to a first-order project is that the basic claim of categoricity – 'there's an isomorphism between N_1 and N_2'– can't be formulated in our first-order language.
- It is possible, however, to devise a first-order formula, $\varphi(x, y)$, that determines a functional relation of the desired form.
- Devising an appropriate φ involves coding.

- Let $\psi(x,u,v)$ say that x, using $+$ and $\cdot$, codes an initial segment I_1 of N_1 ending with u, an initial segment I_2 of N_2 ending with v, and a function $f : I_1 \to I_2$ such that $f(0_1) = 0_2$, $f(z +_1 1_1) = f(z) +_2 1_2$ for all $z \in I_1 \restriction \{u\}$, and $f(u) = v$.
- Let $\varphi(u,v)$ be $\exists x \psi(x,u,v)$.
- Let $\mathrm{ISO}_\varphi(N_1,N_2)$ be the first-order formula which says that φ is a bijection between N_1 and N_2, and for all $x,y \in N_1$:

$$\begin{array}{l} F(0_1) = 0_2 \wedge \\ F(1_1) = 1_2 \wedge \\ F(x +_1 y) = F(x) +_2 F(y) \wedge \\ F(x \cdot_1 y) = F(x) \cdot_2 F(y), \end{array} \tag{1}$$

where $F(x)$ abbreviates the unique $y \in N_2$ such that $\varphi(x,y)$.

With φ as above we obtain:

Theorem ([Vää21])

First-order PA is internally categorical in the sense that

$$PA \cup PA_1(N_1) \cup PA_2(N_2) \vdash \mathrm{ISO}_\varphi(N_1, N_2).$$

Parsons [Par90, Par08] emphasises this kind of categoricity but does not accomplish it. He assumes "Skolem's recursive arithmetic", which is of course included in PA. At the same time he explicitly tries to avoid a common framework such as PA here.

Proof.

By induction on $+_1$, we show that F is a function $N_1 \to N_2$ and that it satisfies the conditions (1). We then use induction on $+_2$ to prove that F is onto. These proofs exploit the fact that induction holds for first-order formulas that have any of the symbols $+, \cdot, 0, 1, +_1, \cdot_1, 0_1, 1_1, +_2, \cdot_2, 0_2, 1_2$. Induction on $+$ is used to establish the necessary properties of the coding. □

It's easy to see that the internal categoricity (and thereby the categoricity) of PA^2 follows from the internal categoricity of the first-order PA.

- Another option is to consider two copies of PA with the same domain: PA_1^* says that the axioms of first-order PA hold of $+_1, \cdot_1, 0_1, 1_1$, PA_2^* says the same for $+_2, \cdot_2, 0_2, 1_2$, where both allow induction for first-order formulas in the joint vocabulary $\{+_1, \cdot_1, 0_1, 1_1\} \cup \{+_2, \cdot_2, 0_2, 1_2\}$.
- If AUT_φ says that φ defines a permutation of the domain and that it satisfies the preservation clauses in the definition of $\mathrm{ISO}_\varphi(N_1, N_2)$.

Theorem ([Vää19, Vää21])

First-order Peano arithmetic PA is internally categorical in the sense that

$$PA_1^* \cup PA_2^* \vdash AUT_\varphi.$$

Proof.

As above, taking $\{+, \cdot, 0, 1\}$ be $\{+_1, \cdot_1, 0_1, 1_1\}$. □

Corollary

If ψ is a first-order sentence, then

$$PA_1^* \cup PA_2^* \vdash \psi(+_1, \cdot_1, 0_1, 1_1) \leftrightarrow \psi(+_2, \cdot_2, 0_2, 1_2).$$

Let $\mathrm{PA}^*_{1,n}$ be the conjunction of the first n axioms of PA^*_1 under some natural enumeration.

Theorem (*'Intolerance' of PA*)

There is a natural number n such that if ψ is a first-order sentence, then

$$\vdash (PA^*_{1,n} \to \psi(+_1, \cdot_1, 0_1, 1_1)) \vee (PA^*_{2,n} \to \neg\psi(+_2, \cdot_2, 0_2, 1_2)).$$

How is this different from $\mathrm{PA} \vdash \psi \vee \neg\psi$? The class of all possible models of PA is divided into those which satisfy ψ and those which satisfy $\neg\psi$.

Corollary

If φ is a first-order sentence of the vocabulary $\{\in\}$, then the theory

$$\{PA_1^* \cup PA_2^*, \psi(+_1, \cdot_1, 0_1, 1_1), \neg\psi(+_2, \cdot_2, 0_2, 1_2)\}$$

is deductively inconsistent.

How is this different from the inconsistency of $\mathrm{PA} \cup \{\varphi\} \cup \{\neg\varphi\}$?

What does internal categoricity of PA^1 give us?

- Categoricity failing, internal categoricity is what is left from the categoricity of PA^2.
- It demonstrates that internal categoricity is not (only) a second order phenomenon.

Internal categoricity of first order theories: set theory

- ZF has the vocabulary $\{\in\}$. ZF(E) is the result of replacing $\in$ with a binary relation symbol E.
- E_1 and E_2 new binary, X_1 and X_2 new unary relation symbols, and π a new unary function symbol.
- Let $\varphi^{(X)}(E)$ be $\varphi(E)$ with the first-order quantifiers relativized to X.
- Let $\mathrm{ZF}^{(X_1)}(E_1)$ consist of all $\varphi^{(X_1)}(E_1)$, where $\varphi \in \mathrm{ZF}$, allowing in the separation and replacement schemas formulas from the vocabulary $\{X_1, E_1, X_2, E_2, \pi\}$ with unrestricted (i.e. not relativized to X_1) quantifiers.
- Similarly $\mathrm{ZF}^{(X_2)}(E_2)$.
- Let IO_π says that π is an isomorphism between the ordinals of (X_1, E_1) and the ordinals of (X_2, E_2).
- Let $\mathrm{ISO}_\varphi((X_1, E_1), (X_2, E_2))$ say that $\varphi(x, y)$ defines an isomorphism between (X_1, E_2) and (X_2, E_2).

Theorem

(Internal quasi-categoricity of ZF) There is a first-order formula $\varphi = \varphi(x, y)$ *of set theory such that*

$$ZF \cup ZF^{(X_1)}(E_1) \cup ZF^{(X_2)}(E_2) \cup \{IO_\pi\} \vdash \mathrm{ISO}_\varphi((X_1, E_1), (X_2, E_2)).$$

- The proof is as below, cf. also [Mar18] for an informal version.
- What fundamentally differentiates this theorem from the second-order version is the mechanism by which that the crucial links between the (X_1, E_1) and (X_2, E_2) are forged: in the first-order theorem, the key is allowing the vocabulary of one into the axiom schemas of the other; in the second-order theorem, these specifics are masked in the Comprehension Axioms.
- [McG97] comes close to the above theorem with his “schematic" categoricity theorem.

- As with arithmetic, the ZF in "the background" can be eliminated if we assume that the domains of the two versions of ZF are the same.
- An extra assumption like IO_π is now unnecessary.
- Given a formula $\varphi(x, y)$, let Aut_φ be the first-order sentence which says that $\varphi(x, y)$ defines an automorphism between the binary predicates E_1 and E_2, again assuming the axioms ZF for E_1 and E_2.
- Let $\mathrm{ZF}(E, E')$ be ZF with E as the membership relation, but with separation and replacement schemas allowing formulas from the vocabulary $\{E, E'\}$.

Theorem ([Vää19])

There is a first-order formula $\varphi = \varphi(x, y)$ of set theory such that

$$ZF(E_1, E_2) \cup ZF(E_2, E_1) \vdash \mathrm{Aut}_{\varphi}(E_1, E_2).$$

Proof of the Theorem

We use $\in_i$ to denote E_i.

- Let $\mathrm{tr}_i(x)$ be the formula $\forall t \in_i x \forall w \in_i t(w \in_i x)$. It says that x is transitive in $\in_i$-set theory.
- Let $\mathrm{TC}_i(x)$ be the unique u such that $\mathrm{tr}_i(u) \wedge x \in_i u \wedge \forall v((\mathrm{tr}_i(v) \wedge x \in_i v) \to \forall w \in_i u(w \in_i v))$ (i.e. "u is the $\in_i$-transitive closure of x").
- Let $\varphi(x, y)$ be the formula $\exists f \psi(x, y, f)$, where $\psi(x, y, f)$ is the conjunction of the following formulas (where $f(t)$ and $f(w)$ are understood in the sense of $\in_1$):

Proof of the Theorem

$$\psi(x, y, f):$$

(1) In the sense of $\in_1$, the set f is a function with $\mathrm{TC}_1(x)$ as its domain.

(2) $\forall t \in_1 \mathrm{TC}_1(x)(f(t) \in_2 \mathrm{TC}_2(y))$

(3) $\forall t \in_2 \mathrm{TC}_2(y) \exists w \in_1 \mathrm{TC}_1(x)(t = f(w))$

(4) $\forall t \in_1 \mathrm{TC}_1(x) \forall w \in_1 \mathrm{TC}_1(x)(t \in_1 w \leftrightarrow f(t) \in_2 f(w))$

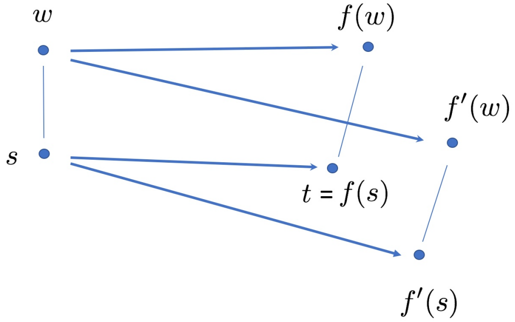
(5) $f(x) = y$

Proof of the Theorem

Lemma

If $\psi(x, y, f)$ and $\psi(x, y, f')$, then $f = f'$.

Proof:



Proof of the Theorem

Lemma

1. *If* $\psi(x, y, f)$ *and* $x' \in_1 x$*, then* $\varphi(x', f(x'))$*.*
2. *If* $\psi(x, y, f)$ *and* $y' \in_2 y$*, then there is* $x' \in_1 x$ *such that* $f(x') = y'$ *and* $\varphi(x', y')$*.*
3. *If* $\varphi(x, y)$ *and* $\varphi(x, y')$*, then* $y = y'$*.*
4. *If* $\varphi(x, y)$ *and* $\varphi(x', y)$*, then* $x = x'$*.*
5. *If* $\varphi(x, y)$ *and* $\varphi(x', y')$*, then* $x' \in_1 x \leftrightarrow y' \in_2 y$*.*

Proof of the Theorem

- Let $\mathrm{On}_1(\mathrm{x})$ be the $\in_1$-formula saying that x is an ordinal i.e. a transitive set of transitive sets, and similarly $\mathrm{On}_2(\mathrm{x})$.
- For $\mathrm{On}_1(\alpha)$ let V^1_α be the α^{th} level of the cumulative hierarchy in the sense of $\in_1$, and similarly V^2_a.

Proof of the Theorem

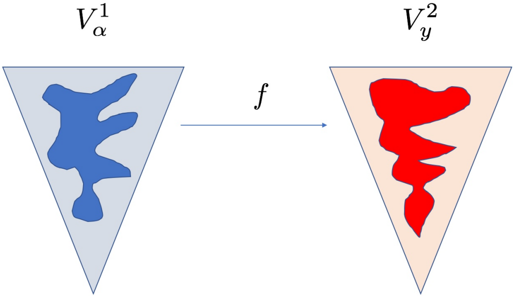
Lemma

1. *If $\varphi(\alpha, y)$, then $\mathrm{On}_1(\alpha)$ if and only if $\mathrm{On}_2(\mathrm{y})$.*
2. *If α is a limit ordinal then so is y i.e. if $\forall u \in_1 \alpha \exists v \in_1 \alpha(u \in_1 v)$, then $\forall u \in_2 y \exists v \in_2 y(u \in_2 v)$.*
3. *Also vice versa.*

Proof of the Theorem

Lemma

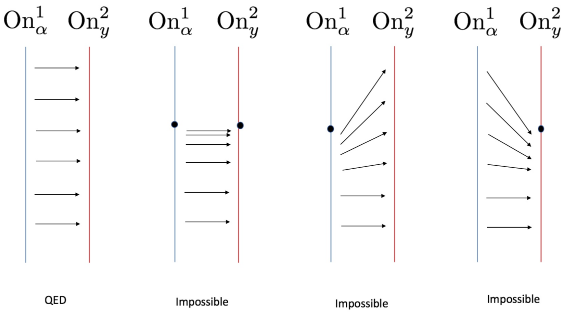
Suppose $\psi(\alpha, y, f)$. If $\mathrm{On}_1(\alpha)$ (or equivalently $\mathrm{On}_2(y)$), then there is $\bar{f} \supseteq f$ such that $\psi(V^1_\alpha, V^2_y, \bar{f})$.



Proof of the Theorem

Lemma

$\forall x \exists y \varphi(x, y)$ *and* $\forall y \exists x \varphi(x, y)$.



- Note that $(M, \in_1)$ and $(M, \in_2)$ can be models of $V = L$, $V \neq L$, CH, $\neg CH$, even of $\neg Con(ZF)$.
- It is easy to construct such pairs of models using classical methods of Gödel and Cohen.
- Not all of them can be models of second order set theory.

Continuum Hypothesis (CH)

- What if $(M, \in_1) \models CH$ and $(M, \in_2) \models \neg CH$?
- Then either $(M, \in_1)$ or $(M, \in_2)$ does not satisfy the Separation Schema or the Replacement Schema if formulas are allowed to mention the other membership-relation.

As before, we can argue:

Corollary

If φ is a first-order sentence in the vocabulary $\{\in\}$, then

$$ZF(E_1, E_2) \cup ZF(E_2, E_1) \vdash \varphi(E_1) \leftrightarrow \varphi(E_2).$$

Let $\mathrm{ZF}_n(E, E')$ be the conjunction of the first n axioms of $\mathrm{ZF}(E, E')$ under some natural enumeration.

Theorem (*'Intolerance' of ZF*)

There is a natural number n such that if φ is a first-order sentence in the vocabulary $\{\in\}$, then

$$\vdash (ZF_n(E_1, E_2) \rightarrow \varphi(E_1)) \vee (ZF_n(E_2, E_1) \rightarrow \neg\varphi(E_2)).$$

How is this different from $\mathrm{ZF} \vdash \varphi \vee \neg\varphi$? The class of all possible models of ZF is divided into those which satisfy φ and those which satisfy $\neg\varphi$.

Corollary

If φ is a first-order sentence of the vocabulary $\{\in\}$, then the theory

$$\{ZF(E_1, E_2), ZF(E_2, E_1), \varphi(E_1), \neg\varphi(E_2)\} \tag{2}$$

is deductively inconsistent.

How is this different from the inconsistency of $\mathrm{ZF} \cup \{\varphi\} \cup \{\neg\varphi\}$?

What does internal (quasi-)categoricity of ZF give us?

- (Quasi-)categoricity failing, internal (quasi-) categoricity is what is left.
- Internal (quasi-)categoricity is not (only) a second order phenomenon.
- A strong robustness result for set theory.
- The model cannot be changed "internally".
- To get non-isomorphic models one has to go "outside" the model.
- But going "outside" raises the potential of an infinite regress of meta theories.

Recap: The importance of internal categoricity in SO

- A strong form of categoricity.
- Independent of set theoretical background.

Recap: The importance of internal categoricity in FO

- A weak form of categoricity.
- An echo of SO axiomatizations.
- Dilutes the first order/second order distinction.
- An outcome that was thought to require second-order resources—namely, categoricity theorems—can actually be achieved by suitable first-order means. What seems to be crucial is a link between the languages of the two relevant models.

A note

There is a connection between internal categoricity and bi-interpretability, brought to my attention recently by Ali Enayat.

Freire and Hamkins: Bi-interpretation in weak set theories. (English summary) J. Symb. Log. 86 (2021), no. 2, 609—634.

Enayat: Variations on a Visserian theme. A tribute to Albert Visser, 99—110, Tributes, 30, Coll. Publ., [London], 2016.

Summary

- Second order logic as the (language for the) foundation of mathematics is advocated because of its power to characterize structures up to isomorphism.
- But full second order logic depends heavily on set theory, which is itself an alternative language for the foundation of mathematics. If you try to avoid set theory, you should not use it as your metatheory.
- Categoricity does not require full second order logic, contrary to common belief. Henkin second order logic is enough for (the stronger) internal categoricity.
- Even first order theories manifest internal categoricity.

Thank you!

Tim Button and Sean Walsh.
Philosophy and model theory.
Oxford University Press, Oxford, 2018.
With a historical appendix by Wilfrid Hodges.

Solomon Feferman.
Infinitary properties, local functors, and systems of ordinal functions.
In *Conference in Mathematical Logic—London '70 (Proc. Conf., Bedford Coll., London, 1970)*, pages 63–97. Lecture Notes in Math., Vol. 255, 1972.

Geoffrey Hellman.
Mathematics without numbers.
The Clarendon Press, Oxford University Press, New York, 1989.
Towards a modal-structural interpretation.

Shaughan Lavine.
Skolem was wrong.
Unpublished, 1999.

Donald Martin.
Completeness or incompleteness of basic mathematical concepts (draft).
http://www.math.ucla.edu/ dam/booketc/efi.pdf, 2018.

Vann McGee.
How we learn mathematical language.
Philosophical Review, 106(1):35–68, 1997.

Charles Parsons.
The uniqueness of the natural numbers.
Iyyun: The Jerusalem Philosophical Quarterly, 39:13–44, 1990.

Charles Parsons.
Mathematical thought and its objects.
Cambridge University Press, Cambridge, 2008.

Stewart Shapiro.
Foundations without foundationalism, volume 17 of *Oxford Logic Guides*.
The Clarendon Press, Oxford University Press, New York, 1991.
A case for second-order logic, Oxford Science Publications.

Jouko Väänänen.
An extension of a theorem of Zermelo.
Bull. Symb. Log., 25(2):208–212, 2019.

Jouko Väänänen.
Tracing internal categoricity.

Theoria, 87:986–1000, 2021.

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

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