A Logical Framework Perspective on Conservativity

Florian Rabe

University Erlangen-Nuremberg

August 2024

What is a Conservative Extension?

Intuition

Concepts

- theory: list of declarations in some language
- extension E of T: like T but with additional declarations
- conservative: E is kind of the same as T

Example:

T = natural numbers with 0, successor

$$\mathbb{N} : \mathtt{type} \quad 0 : \mathbb{N} \quad s : \mathbb{N} \to \mathbb{N} \quad \text{Peano axioms}$$

E = extension with addition

$$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \quad \forall x.x + 0 = x \quad \forall x.\forall n.x + s(n) = s(x + n)$$

Kind of the same?

- same syntax?
- same semantics? conservativity in model theory
- same theorems? condervativity in proof theory

Common Patterns for Conservative Extensions

- new definition — side note: not usually considered in logic

$$plusTwo := \lambda x.s(s(x))$$

- new type

$$foo : \mathtt{type}$$

- new term of new type

$$bar : \mathbb{N} \to foo$$

- new terms of existing types — if $\mathbb{N}$ is non-empty

$$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$

- …with certain axioms — encoding a definition

$$\forall x.x + 0 = x, \quad \forall x.\forall n.x + s(n) = s(x + n)$$

- Skolemization, e.g., of $\forall n.\exists x.n = 0 \vee s(x) = n$

$$\mathtt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \vee s(\mathtt{pred}(n)) = n$$

Common Patterns for Conservative Extensions

- new definition side note: not usually considered in logic

$$plusTwo := \lambda x.s(s(x))$$

- new type

$$foo : \texttt{type}$$

- new term of new type

$$bar : \mathbb{N} \to foo$$

- new terms of existing types if $\mathbb{N}$ is non-empty

$$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$

- ...with certain axioms encoding a definition

$$\forall x.x + 0 = x, \quad \forall x.\forall n.x + s(n) = s(x + n)$$

- Skolemization, e.g., of $\forall n.\exists x.n = 0 \vee s(x) = n$

$$\texttt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \vee s(\texttt{pred}(n)) = n$$

Common Patterns for Conservative Extensions

- new definition side note: not usually considered in logic

$$plusTwo := \lambda x.s(s(x))$$

- new type

$$foo : \mathtt{type}$$

- new term of new type

$$bar : \mathbb{N} \rightarrow foo$$

- new terms of existing types if $\mathbb{N}$ is non-empty

$$+ : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$$

- ...with certain axioms encoding a definition

$$\forall x.x + 0 = x, \quad \forall x.\forall n.x + s(n) = s(x + n)$$

- Skolemization, e.g., of $\forall n.\exists x.n = 0 \vee s(x) = n$

$$\mathtt{pred} : \mathbb{N} \rightarrow \mathbb{N},\ \forall n.n = 0 \vee s(\mathtt{pred}(n)) = n$$

Common Patterns for Conservative Extensions

- new definition — side note: not usually considered in logic

$$plusTwo := \lambda x.s(s(x))$$

- new type

$$foo : \texttt{type}$$

- new term of new type

$$bar : \mathbb{N} \to foo$$

- new terms of existing types — if $\mathbb{N}$ is non-empty

$$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$

- …with certain axioms — encoding a definition

$$\forall x.x + 0 = x, \quad \forall x.\forall n.x + s(n) = s(x + n)$$

- Skolemization, e.g., of $\forall n.\exists x.n = 0 \vee s(x) = n$

$$\texttt{pred} : \mathbb{N} \to \mathbb{N}, \ \forall n.n = 0 \vee s(\texttt{pred}(n)) = n$$

Common Patterns for Conservative Extensions

- new definition side note: not usually considered in logic

$$plusTwo := \lambda x.s(s(x))$$

- new type

$$foo : \texttt{type}$$

- new term of new type

$$bar : \mathbb{N} \to foo$$

- new terms of existing types if $\mathbb{N}$ is non-empty

$$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$

- ...with certain axioms encoding a definition

$$\forall x.x + 0 = x, \quad \forall x.\forall n.x + s(n) = s(x + n)$$

- Skolemization, e.g., of $\forall n.\exists x.n = 0 \vee s(x) = n$

$$\texttt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \vee s(\texttt{pred}(n)) = n$$

Common Patterns for Conservative Extensions

- new definition — side note: not usually considered in logic

$$plusTwo := \lambda x.s(s(x))$$

- new type

$$foo : \mathtt{type}$$

- new term of new type

$$bar : \mathbb{N} \to foo$$

- new terms of existing types — if $\mathbb{N}$ is non-empty

$$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$$

- ...with certain axioms — encoding a definition

$$\forall x.x + 0 = x, \quad \forall x.\forall n.x + s(n) = s(x + n)$$

- Skolemization, e.g., of $\forall n.\exists x.n = 0 \lor s(x) = n$

$$\mathtt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \lor s(\mathtt{pred}(n)) = n$$

Formal Definitions

Competing definitions

$T \hookrightarrow E$ is conservative

- **syntactically** (SC)

  everything added in E can already be defined in T
- **proof-theoretically** (PC)

  E can't prove anything new about T
- **model-theoretically** (MC)

  T-models can be extended to E-models

definitions are from the literature, names are mine

preview:

$$SC \Rightarrow \ldots \Rightarrow MC \Rightarrow \ldots \Rightarrow PC$$

Competing definitions

$T \hookrightarrow E$ is conservative

- **syntactically** (SC)

  everything added in E can already be defined in T

- **proof-theoretically** (PC)

  E can't prove anything new about T

- **model-theoretically** (MC)

  T-models can be extended to E-models

definitions are from the literature, names are mine

preview:

$$\text{SC} \Rightarrow \ldots \Rightarrow \text{MC} \Rightarrow \ldots \Rightarrow \text{PC}$$

Definition 1: Syntactic Conservativity

Setting

- formal system with theories and morphism
  e.g., FOL, HOL, CoC, ZFC
- theory extension $T \hookrightarrow E$ is inclusion morphism

Syntactically conservative (SC) iff

- there is morphism $d : E \to T$ that is identity on T
- in other words: T can define all constants added in E

Example

- T=Peano, E adds $1 : \mathbb{N}$
- define d by $d(1) = s(0)$

Are the Common Patterns SC?

Some patterns are always SC

- new definition $plusTwo := \lambda x : \mathbb{N}.s(s(x))$

$$d(plusTwo) = \text{its definiens}$$

- new type foo : type

$$d(foo) = \text{any type}$$

- new term $bar : \mathbb{N} \to foo$

$$d(bar) = \lambda n : \mathbb{N}.\text{any term of type } d(foo)$$

- new term of existing type $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

$$d(+) = \lambda m : \mathbb{N}, n : \mathbb{N}.\text{any term of type } \mathbb{N}$$

Adding terms with axioms is often not SC

- $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and axioms
- $\mathtt{pred} : \mathbb{N} \to \mathbb{N}$ and axiom

no way to define $d(+)$ or $d(\mathtt{pred})$ that satisfies the axioms

Are the Common Patterns SC?

Some patterns are always SC

- new definition $plusTwo := \lambda x : \mathbb{N}.s(s(x))$

  $$d(plusTwo) = \text{its definiens}$$

- new type foo : type

  $$d(foo) = \text{any type}$$

- new term $bar : \mathbb{N} \to foo$

  $$d(bar) = \lambda n : \mathbb{N}.\text{any term of type } d(foo)$$

- new term of existing type $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

  $$d(+) = \lambda m : \mathbb{N}, n : \mathbb{N}.\text{any term of type } \mathbb{N}$$

Adding terms with axioms is often not SC

- $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and axioms
- $\mathtt{pred} : \mathbb{N} \to \mathbb{N}$ and axiom

no way to define $d(+)$ or $d(\mathtt{pred})$ that satisfies the axioms

Are the Common Patterns SC?

Some patterns are always SC

- new definition $plusTwo := \lambda x : \mathbb{N}.s(s(x))$

$$d(plusTwo) = \text{its definiens}$$

- new type $foo : \text{type}$

$$d(foo) = \text{any type}$$

- new term $bar : \mathbb{N} \to foo$

$$d(bar) = \lambda n : \mathbb{N}.\text{any term of type } d(foo)$$

- new term of existing type $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

$$d(+) = \lambda m : \mathbb{N}, n : \mathbb{N}.\text{any term of type } \mathbb{N}$$

Adding terms with axioms is often not SC

- $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and axioms
- $\mathtt{pred} : \mathbb{N} \to \mathbb{N}$ and axiom

no way to define $d(+)$ or $d(\mathtt{pred})$ that satisfies the axioms

Are the Common Patterns SC?

Some patterns are always SC

- new definition $plusTwo := \lambda x : \mathbb{N}.s(s(x))$

$$d(plusTwo) = \text{its definiens}$$

- new type foo : type

$$d(foo) = \text{any type}$$

- new term $bar : \mathbb{N} \to foo$

$$d(bar) = \lambda n : \mathbb{N}.\text{any term of type } d(foo)$$

- new term of existing type $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

$$d(+) = \lambda m : \mathbb{N}, n : \mathbb{N}.\text{any term of type } \mathbb{N}$$

Adding terms with axioms is often not SC

- $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and axioms
- $\mathtt{pred} : \mathbb{N} \to \mathbb{N}$ and axiom

no way to define $d(+)$ or $d(\mathtt{pred})$ that satisfies the axioms

Are the Common Patterns SC?

Some patterns are always SC

- new definition $plusTwo := \lambda x : \mathbb{N}.s(s(x))$

$$d(plusTwo) = \text{its definiens}$$

- new type foo : type

$$d(foo) = \text{any type}$$

- new term $bar : \mathbb{N} \to foo$

$$d(bar) = \lambda n : \mathbb{N}.\text{any term of type } d(foo)$$

- new term of existing type $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

$$d(+) = \lambda m : \mathbb{N}, n : \mathbb{N}.\text{any term of type } \mathbb{N}$$

Adding terms with axioms is often not SC

- $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ and axioms
- $\texttt{pred} : \mathbb{N} \to \mathbb{N}$ and axiom

no way to define $d(+)$ or $d(\texttt{pred})$ that satisfies the axioms

Definition 2: Proof-Theoretic Conservativity

Setting

- Inclusion morphisms $T \hookrightarrow E$ as before
- Language has propositions $F : \mathtt{prop}$ and truth judgment $\vdash F$

i.e., it's a logic

Proof-theoretically conservative (PC) iff

- informally:

  any E-provable T-proposition is also T-provable
- formally:

  for $\vdash_T F : \mathtt{prop}$, if $\vdash_E F$, then also $\vdash_T F$
- intuitively:

  E doesn't change T-provability

Note that $\vdash_T F$ always implies $\vdash_E F$
independent of conservativity

Are the Common Patterns PC?

Theorem: SC already implies PC

Proof: Assume $d : E \to T$ and T-formula F.
Any E-proof P of F yields a T-proof $d(P)$ of $d(F) = F$.

Our other examples are PC as well:

- new terms with defining axioms

  $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \quad \forall x.x + 0 = x \quad \forall x.x \forall n.x + s(n) = s(x + n)$

  PC: axioms carefully chosen to rewrite new terms into old ones
- Skolemization of $\forall n.\exists x.n = 0 \vee s(x) = n$

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \vee s(\mathtt{pred}(n)) = n$$

  PC: equivalent to soundness of Skolemization

Are the Common Patterns PC?

Theorem: SC already implies PC

Proof: Assume $d : E \to T$ and T-formula F.
Any E-proof P of F yields a T-proof $d(P)$ of $d(F) = F$.

Our other examples are PC as well:

- new terms with defining axioms

  $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \quad \forall x.x + 0 = x \quad \forall x.x\forall n.x + s(n) = s(x + n)$

  PC: axioms carefully chosen to rewrite new terms into old ones

- Skolemization of $\forall n.\exists x.n = 0 \vee s(x) = n$

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \vee s(\mathtt{pred}(n)) = n$$

  PC: equivalent to soundness of Skolemization

Are the Common Patterns PC?

Theorem: SC already implies PC

Proof: Assume $d : E \to T$ and T-formula F.
Any E-proof P of F yields a T-proof $d(P)$ of $d(F) = F$.

Our other examples are PC as well:

- new terms with defining axioms

  $$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \quad \forall x.x + 0 = x \quad \forall x.x \forall n.x + s(n) = s(x + n)$$

  PC: axioms carefully chosen to rewrite new terms into old ones

- Skolemization of $\forall n.\exists x.n = 0 \vee s(x) = n$

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \vee s(\mathtt{pred}(n)) = n$$

  PC: equivalent to soundness of Skolemization

Are the Common Patterns PC?

Theorem: SC already implies PC
Proof: Assume $d : E \to T$ and T-formula F.
Any E-proof P of F yields a T-proof $d(P)$ of $d(F) = F$.

Our other examples are PC as well:

- new terms with defining axioms

  $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \quad \forall x.x + 0 = x \quad \forall x.x\forall n.x + s(n) = s(x + n)$

  PC: axioms carefully chosen to rewrite new terms into old ones

- Skolemization of $\forall n.\exists x.n = 0 \vee s(x) = n$

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N},\ \forall n.n = 0 \vee s(\mathtt{pred}(n)) = n$$

  PC: equivalent to soundness of Skolemization

Definition 3: Model-Theoretic Conservativity

Setting

- Inclusion morphisms $T \hookrightarrow E$ as before
- Logic has sound+complete proof and model theory
- Forgetful functors $|_T : \mathbf{Mod}(E) \to \mathbf{Mod}(T)$

e.g., models are interpretation function; $|_T$ restricts to T-syntax

Model-theoretically conservative (MC) iff

- Any T-model $M \in \mathbf{Mod}(T)$
  can be extended to an E-model $N \in \mathbf{Mod}(E)$
  such that $N|_T = M$.
- Intuitively: All old models can be extended to the extended syntax.

in other words: forgetful functor is surjective

Are the Common Patterns MC?

Theorem: SC implies MC

Proof: Assume $d : T \to E$ and a T-model.
"Composing" d and T yields an E-model.

Our other examples are (usually) MC as well:

- new terms with defining axioms

  $$+ : \mathbb{N}\times\mathbb{N} \to \mathbb{N} \quad \forall x.x+0 = x \quad \forall x.\forall n : \mathbb{N}.x+s(n) = s(x+n)$$

  MC: extend Peano models with addition function

- Skolemization of provable $\forall\exists$ formula

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N} \quad \forall n.n = 0 \vee s(\mathtt{pred}(n)) = n$$

  MC: extend Peano models with predecessor function

Are the Common Patterns MC?

Theorem: SC implies MC

Proof: Assume $d : T \to E$ and a T-model.
"Composing" d and T yields an E-model.

Our other examples are (usually) MC as well:

- new terms with defining axioms

  $+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \quad \forall x.x+0 = x \quad \forall x.\forall n : \mathbb{N}.x+s(n) = s(x+n)$

  MC: extend Peano models with addition function

- Skolemization of provable $\forall\exists$ formula

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N} \quad \forall n.n = 0 \lor s(\mathtt{pred}(n)) = n$$

  MC: extend Peano models with predecessor function

Are the Common Patterns MC?

Theorem: SC implies MC

Proof: Assume $d : T \to E$ and a T-model.
"Composing" d and T yields an E-model.

Our other examples are (usually) MC as well:

- new terms with defining axioms

  $$+ : \mathbb{N}\times\mathbb{N} \to \mathbb{N} \quad \forall x.x+0 = x \quad \forall x.\forall n : \mathbb{N}.x+s(n) = s(x+n)$$

  MC: extend Peano models with addition function

- Skolemization of provable ∀∃ formula

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N} \quad \forall n.n = 0 \lor s(\mathtt{pred}(n)) = n$$

  MC: extend Peano models with predecessor function

Are the Common Patterns MC?

Theorem: SC implies MC

Proof: Assume $d : T \to E$ and a T-model.
"Composing" d and T yields an E-model.

Our other examples are (usually) MC as well:

- new terms with defining axioms

  $$+ : \mathbb{N}\times\mathbb{N} \to \mathbb{N} \quad \forall x.x+0 = x \quad \forall x.\forall n : \mathbb{N}.x+s(n) = s(x+n)$$

  MC: extend Peano models with addition function

- Skolemization of provable $\forall\exists$ formula

  $$\mathtt{pred} : \mathbb{N} \to \mathbb{N} \quad \forall n.n = 0 \lor s(\mathtt{pred}(n)) = n$$

  MC: extend Peano models with predecessor function

MC vs. PC

Proved so far

- SC implies PC
- SC implies MC

Theorem: MC implies PC

Proof: Assume E-provable T-proposition F

- by soundness: F holds in all E-models
- by MC: F holds in all T-models
- by completeness: F is T-provable

Question: Does PC imply MC?

MC vs. PC

Proved so far

- SC implies PC
- SC implies MC

Theorem: MC implies PC

Proof: Assume E-provable T-proposition F

- by soundness: F holds in all E-models
- by MC: F holds in all T-models
- by completeness: F is T-provable

Question: Does PC imply MC?

Degrees of MC between SC and PC

Example

- first-order logic with usual proof theory
- theory T

$$f : a \to b \qquad \text{axiom that } f \text{ surjective}$$

- extension E

$$r : b \to a \qquad \text{axiom } \forall x.r(f(x)) = x$$

- $T \hookrightarrow E$ is PC, but not SC

Questions: Is it MC?

Degrees of MC between SC and PC

Example

- first-order logic with usual proof theory
- theory T

$$f : a \to b \quad \text{axiom that } f \text{ surjective}$$

- extension E

$$r : b \to a \quad \text{axiom } \forall x.r(f(x)) = x$$

- $T \hookrightarrow E$ is PC, but not SC

Questions: Is it MC?

- using ZF: no
- using ZFC: yes

Results

The Conservativity Spectrum

SC $\Rightarrow$ MC for weaker sem. $\Rightarrow$ MC for stronger sem. $\Rightarrow$ PC

SC

- proved using only the syntax — morphism d is syntactic witness
- strongest reasonable definition

PC

- proved using arbitrary extra-logical means
- weakest reasonable definition

MC

- proved using model theory — stronger than syntax, but still limited
- stronger model theory = more conservative extensions

Background: Logics in the MMT Framework

Logics as theories, semantics as translations

- no distinction between logic translation and semantics

$$FOL \to HOL \to ZF \hookrightarrow ZFC \to CoC \ldots$$

- semantics = translation $sem : Logic \to Foundation$

Logic = syntax+calculus
Foundation = e.g., ZFC, CoC, ...
sem = interpretation function + soundness proof

Example: Standard semantics of FOL

- MMT-theory for ZFC with set, $\in$, ...
- morphism FOL$\to$ ZFC maps
  - types to sets
  - terms to elements
  - formulas to Booleans $\{0, 1\}$
  - theorems to Boolean 1

Unifying Definition

Given $sem : Logic \to Foundation$

translates $T \hookrightarrow E$ to $sem(T) \hookrightarrow sem(E)$

Define: $T \hookrightarrow E$ *sem*-conservative if

exists retraction $d : sem(E) \to sem(T)$

must be identity on $sem(T)$

SC and PC as special cases

- SC: $sem =$ identity
  syntax as initial semantics: everything interpreted as itself
- MC: *sem* is model theory of the logic   gradual refinement of syntax
- PC: models are maximal consistent theories
  proof theory as terminal semantics

Theorem: conservativity preserved along translations

Conservativity of Morphisms

So far

- conservativity of an extension $T \hookrightarrow E$
- the typical case considered by the community
- but actually: needlessly specific

Generalization

- conservativity for arbitrary morphisms $m : T \to E$
  - SC: morphism $d : E \to T$ such that $m; d = id_T$
  - PC: if $m(F)$ is E-provable, then F is T-provable
  - MC: T-models induce E-models by factoring through m
- all results carry over the general case

definitions/proofs actually slightly easier

Conservativity of Language Extensions

So far

- conservativity of theory extensions $E \hookrightarrow T$ in some logic L
- the typical case formalized by the community

Generalization

- conservativity of logic extension $L \hookrightarrow M$
  in MMT: L, M represented as theories of the logical framework
- essentially: SC=derivable, PC=admissible — all results carry over
- examples:
  - sequent calculus extended with cut rule
  - λ-calculus extended with product types
  - set theory with description operator

Completeness is Conservativity of Semantics

Putting both generalizations together

- given: semantics as language morphism $sem : Logic \to Foundation$
- define: SC and PC for *sem*

Main theorem

- logic+semantics is complete iff *sem* is PC
- proved in generality in the MMT framework

Conclusion

Key Takeaways

- logics are theories, semantics are morphisms in logical frameworks
- composing morphisms yields increasingly refined semantics
- conservativity spectrum of semantics
  - SC and PC as extreme points
  - conservativity preserved along refinement
- generalized conservativity to arbitrary morphisms in the framework
  intra- or inter-logical
- Conservativity = Completeness

What's in the paper?

- formal definition of arbitrary logic
- logic-independent in MMT framework
- definitions and theorems for arbitrary logics