Free Logic in HOL
(On Cats & Alligators and Why Everything is Defined)

Christoph Benzmüller[1] — FU Berlin & Visiting Scholar at Stanford U.
Dana Scott — Visiting Scholar at UC Berkeley

ICMS 2016, Berlin, July 11, 2016

[1] Supported by DFG Heisenberg Fellowship BE 2501/9-1/2

Talk Outline

A HOL as a Universal (Meta-)Logic via Semantic Embeddings

B Free Logic in HOL

C Application

- Categories and Allegories – Textbook by Freyd & Scedrov, 1990
- ATPs revealed a kind of inconsistency in this book:

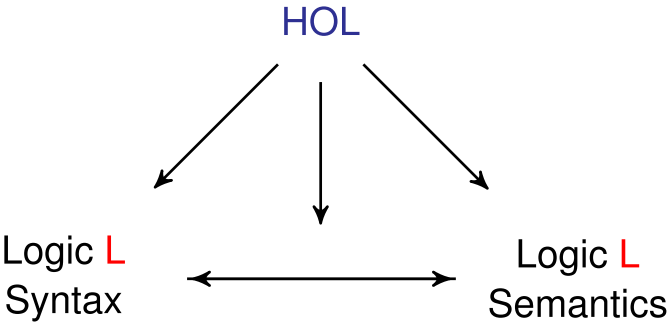
> "Either all morphisms exist or Freyd's axioms are inconsistent"

D Remark on Verified Publications

E Conclusion

Part A:
HOL as a Universal (Meta-)Logic via Semantic Embeddings

HOL as a Universal (Meta-)Logic via Semantic Embeddings

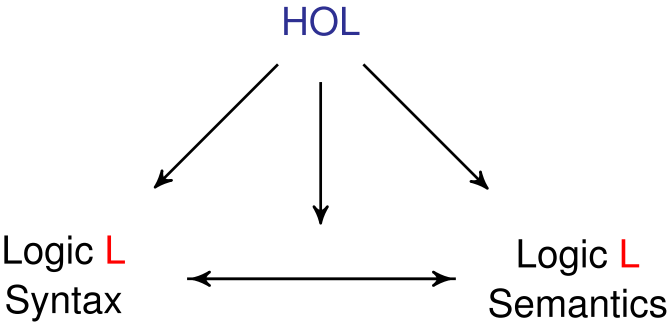


Examples for L we have already studied:

Modal Logics, Conditional Logics, Intuitionistic Logics, Access Control Logics, Nominal Logics, Multivalued Logics (SIXTEEN), Logics based on Neighbourhood Semantics, (Mathematical) Fuzzy Logics, Paraconsistent Logics, Free Logic . . .

Works also for (first-order & higher-order) quantifiers

HOL as a Universal (Meta-)Logic via Semantic Embeddings



Examples for L we have already studied:
Modal Logics, Conditional Logics, Intuitionistic Logics, Access Control Logics, Nominal Logics, Multivalued Logics (SIXTEEN), Logics based on Neighbourhood Semantics, (Mathematical) Fuzzy Logics, Paraconsistent Logics, Free Logic . . .

Works also for (first-order & higher-order) quantifiers

Embedding Approach — Idea

HOL (meta-logic) φ ::=

Your-logic (object-logic) ψ ::=

Embedding of in

=

=

=

=

Embedding of meta-logical notions on in

valid =

satisfiable =

... =

Pass this set of equations to a higher-order automated theorem prover

Classical Higher-Order Logic (HOL) — Church's Simple Type Theory (1940)

Simple Types $\alpha ::= o \mid \iota \mid \mu \mid \alpha_1 \to \alpha_2$

HOL $s,t ::= c_\alpha \mid x_\alpha \mid (\lambda x_\alpha s_\beta)_{\alpha\to\beta} \mid (s_{\alpha\to\beta}\, t_\alpha)_\beta \mid$
$(\neg_{o\to o}\, s_o)_o \mid (s_o \vee_{o\to o\to o} t_o)_o \mid (\forall_{(\alpha\to o)\to o}(\lambda x_\alpha t_o))_o$

(note: binder notation $\forall x_\alpha t_o$ as syntactic sugar for $\forall_{(\alpha\to o)\to o}(\lambda x_\alpha t_o)$)

HOL with Henkin semantics is (meanwhile) well understood

- Origin [Church,JSymbLog,1940]
- Henkin semantics [Henkin,JSymb.Log,1950]
  [Andrews, JSymbLog,1971,1972]
- Extens./Intens. [BenzmüllerEtAl,JSymbLog,2004]
  [Muskens,JSymbLog,2007]

Sound and complete provers do exists

- interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, ...
- automated: TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, ...

Classical Higher-Order Logic (HOL) — Church's Simple Type Theory (1940)

Simple Types $\alpha ::= o \mid \iota \mid \mu \mid \alpha_1 \rightarrow \alpha_2$

HOL $s, t ::= \boldsymbol{c}_\alpha \mid x_\alpha \mid (\lambda x_\alpha s_\beta)_{\alpha\rightarrow\beta} \mid (s_{\alpha\rightarrow\beta}\, t_\alpha)_\beta \mid$

$$(\neg_{o\rightarrow o}\ s_o)_o \mid (s_o \vee_{o\rightarrow o\rightarrow o} t_o)_o \mid (\forall_{(\alpha\rightarrow o)\rightarrow o}(\lambda x_\alpha t_o))_o$$

(note: binder notation $\forall x_\alpha t_o$ as syntactic sugar for $\forall_{(\alpha\rightarrow o)\rightarrow o}(\lambda x_\alpha t_o)$)

HOL with Henkin semantics is (meanwhile) well understood

- Origin [Church,JSymbLog,1940]
- Henkin semantics [Henkin,JSymb.Log,1950] [Andrews, JSymbLog,1971,1972]
- Extens./Intens. [BenzmüllerEtAl,JSymbLog,2004] [Muskens,JSymbLog,2007]

Sound and complete provers do exists

- interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, . . .
- automated: TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, . . .

Classical Higher-Order Logic (HOL) — Church's Simple Type Theory (1940)

Simple Types $\alpha ::= o \mid \iota \mid \mu \mid \alpha_1 \rightarrow \alpha_2$

HOL $s, t ::= \boldsymbol{c}_\alpha \mid x_\alpha \mid (\lambda x_\alpha s_\beta)_{\alpha\rightarrow\beta} \mid (s_{\alpha\rightarrow\beta}\, t_\alpha)_\beta \mid$

$(\neg_{o\rightarrow o}\, s_o)_o \mid (s_o \vee_{o\rightarrow o\rightarrow o} t_o)_o \mid (\forall_{(\alpha\rightarrow o)\rightarrow o}(\lambda x_\alpha t_o))_o$

(note: binder notation $\forall x_\alpha t_o$ as syntactic sugar for $\forall_{(\alpha\rightarrow o)\rightarrow o}(\lambda x_\alpha t_o)$)

HOL with Henkin semantics is (meanwhile) well understood

- Origin [Church,JSymbLog,1940]
- Henkin semantics [Henkin,JSymb.Log,1950]
  [Andrews, JSymbLog,1971,1972]
- Extens./Intens. [BenzmüllerEtAl,JSymbLog,2004]
  [Muskens,JSymbLog,2007]

Sound and complete provers do exists

- interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, ...
- automated: TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, ...

Classical Higher-Order Logic (HOL) — Church's Simple Type Theory (1940)

Simple Types $\alpha ::= o \mid \iota \mid \mu \mid \alpha_1 \rightarrow \alpha_2$

HOL $s, t ::= \boldsymbol{c}_\alpha \mid x_\alpha \mid (\lambda x_\alpha s_\beta)_{\alpha\rightarrow\beta} \mid (s_{\alpha\rightarrow\beta}\, t_\alpha)_\beta \mid$

$$(\neg_{o\rightarrow o}\; s_o)_o \mid (s_o \vee_{o\rightarrow o\rightarrow o} t_o)_o \mid (\forall_{(\alpha\rightarrow o)\rightarrow o}(\lambda x_\alpha t_o))_o$$

(note: binder notation $\forall x_\alpha t_o$ as syntactic sugar for $\forall_{(\alpha\rightarrow o)\rightarrow o}(\lambda x_\alpha t_o)$)

HOL with Henkin semantics is (meanwhile) well understood

- Origin [Church,JSymbLog,1940]
- Henkin semantics [Henkin,JSymb.Log,1950] [Andrews, JSymbLog,1971,1972]
- Extens./Intens. [BenzmüllerEtAl,JSymbLog,2004] [Muskens,JSymbLog,2007]

Sound and complete provers do exists

- interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, ...
- automated: TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, ...

Prominent Applications of the Semantic Embedding Approach


Home | Video | Themen | Forum | English | DER SPIEGEL | SPIEGEL TV | Abo | Shop RSS | Mobile | Newsletter

SPIEGEL ONLINE INTERNATIONAL Sign in | Register

Front Page World Europe Germany Business Zeitgeist Newsletter

English Site > Germany > Science > Scientists Use Computer to Mathematically Prove Gödel God Theorem

Holy Logic: Computer Scientists 'Prove' God Exists

By David Knight

picture-alliance/ Imagno/ Wiener Stadt- und Landesbibliothek

Austrian mathematician Kurt Gödel kept his proof of God's existence a secret for decades. Now two scientists say they have proven it mathematically using a computer.

Two scientists have formalized a theorem regarding the existence of God penned by mathematician Kurt Gödel. But the God angle is somewhat of a red herring -- the real step forward is the example it sets of how computers can make scientific progress simpler.

See results reported in

- **Automating Gödel's Ontological Proof of God's Existence with Higher-order Automated Theorem Provers**, ECAI 2014
- **The Inconsistency in Gödel's Ontological Argument: A Success Story for AI in Metaphysics**, IJCAI 2016

Discussion of this work was starting point for collaboration with Dana Scott.

Part B:
Free Logic (Scott, 1967) in HOL

Free Logic: Elegant Approach to Definite Description and Undefinedness

Dana Scott. "Existence and description in formal logic." In: Bertrand Russell: Philosopher of the Century, edited by R. Schoenman. George Allen & Unwin, London, 1967, pp. 181-200. Reprinted with additions in: Philosophical Application of Free Logic, edited by K. Lambert. Oxford Universitry Press, 1991, pp. 28 - 48.

16

DANA SCOTT

Existence and Description in Formal Logic

The problem of what to do with improper descriptive phrases has bothered logicians for a long time. There have been three major suggestions of how to treat descriptions usually associated with the names of Russell, Frege and Hilbert-Bernays. The author does not consider any of these approaches really satisfactory. In many ways Russell's idea is most attractive because of its simplicity. However, on second thought one is saddened to find that the Russellian method of elimination depends heavily on the scope of the elimination.

Previous Approaches (rough sketch)

The present King of France is bald.

.

Russel (first approach)

$bald(\iota x.pKoF(x))$

iff

$(\exists x.pKoF(x)) \wedge (\forall x, y.pKoF((x) \wedge pKoF((y) \rightarrow x = y) \wedge (\forall x.pKoF((x) \rightarrow bald(x))$

Hence, false.

Frege

$\iota x.pKoF(x)$ does not denote; $bald(\iota x.pKoF(x))$ has no truth value.

Hilbert-Bernays

If the existence and uniqueness conditions cannot be proved, then the term $\iota x.pKoF(x)$ cannot be introduced in the language.

Free Logic: Elegant Approach to Definite Description and Undefinedness

Existence and Description in Formal Logic (Dana Scott), 1967

Principle 1: Bound individual variables range over domain $E \subset D$

Principle 2: Domain E may be empty

Principle 3: Values of terms and free variables are in D, not necessarily in E only.

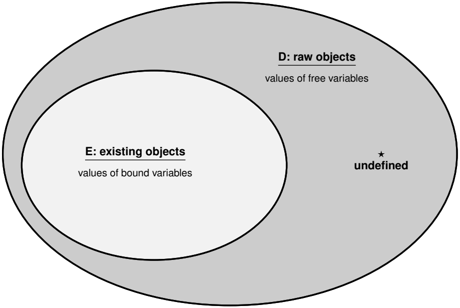
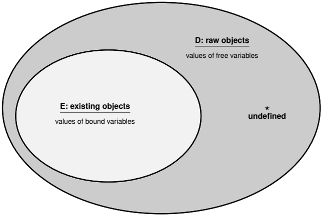


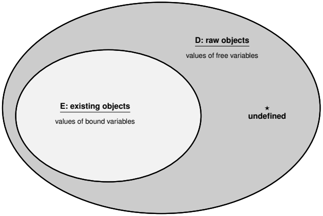
Figure: Illustration of the Semantical Domains of Free Logic

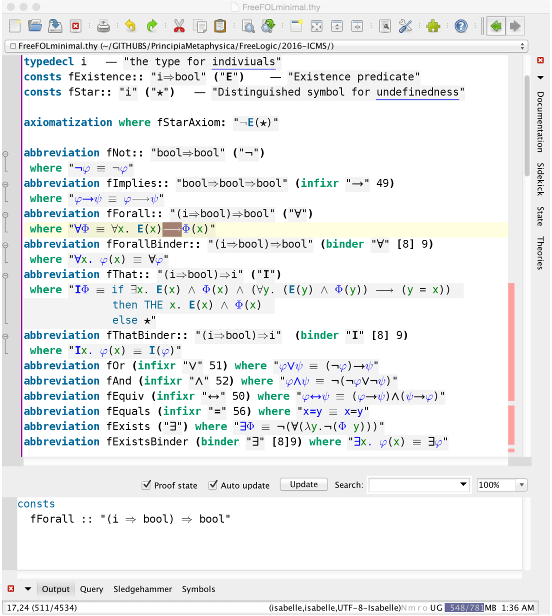
Easy Formalisation in HOL



- Raw domain D_i: type i
- Subdomain E: predicate $E_{i \to o}$
- $\neg$ mapped to $\neg$ and $\longrightarrow$ to $\longrightarrow$
- $\forall x.\varphi(x)$ mapped to $\forall x.E(x) \longrightarrow \varphi(x)$
- Other connectives defined as usual
- $\iota x.\varphi(x)$ mapped to
  $\mathit{if}\ E(x) \wedge \forall y.E(y) \wedge \varphi(y) \longrightarrow y = x$
  $\mathit{then}\ \iota x.E(x) \wedge \varphi(x)$
  $\mathit{else}\ \star$

Free Logic in HOL

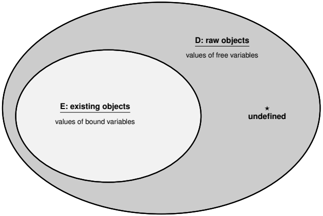




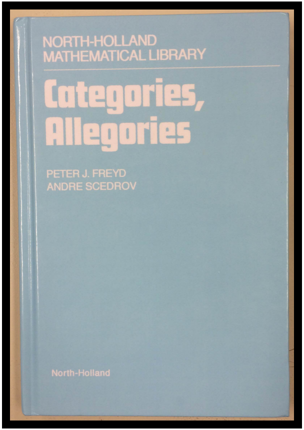

```
typedecl i   — "the type for indiviuals"
consts fExistence:: "i⇒bool" ("E")    — "Existence predicate"
consts fStar:: "i" ("⋆")    — "Distinguished symbol for undefinedness"

axiomatization where fStarAxiom: "¬E(⋆)"

abbreviation fNot:: "bool⇒bool" ("¬")
 where "¬φ ≡ ¬φ"
abbreviation fImplies:: "bool⇒bool⇒bool" (infixr "→" 49)
 where "φ→ψ ≡ φ⟶ψ"
abbreviation fForall:: "(i⇒bool)⇒bool" ("∀")
 where "∀Φ ≡ ∀x. E(x)⟶Φ(x)"
abbreviation fForallBinder:: "(i⇒bool)⇒bool" (binder "∀" [8] 9)
 where "∀x. φ(x) ≡ ∀φ"
abbreviation fThat:: "(i⇒bool)⇒i" ("I")
 where "IΦ ≡ if ∃x. E(x) ∧ Φ(x) ∧ (∀y. (E(y) ∧ Φ(y)) ⟶ (y = x))
              then THE x. E(x) ∧ Φ(x)
              else ⋆"
abbreviation fThatBinder:: "(i⇒bool)⇒i"  (binder "I" [8] 9)
 where "Ix. φ(x) ≡ I(φ)"
abbreviation fOr (infixr "∨" 51) where "φ∨ψ ≡ (¬φ)→ψ"
abbreviation fAnd (infixr "∧" 52) where "φ∧ψ ≡ ¬(¬φ∨¬ψ)"
abbreviation fEquiv (infixr "↔" 50) where "φ↔ψ ≡ (φ→ψ)∧(ψ→φ)"
abbreviation fEquals (infixr "=" 56) where "x=y ≡ x=y"
abbreviation fExists ("∃") where "∃Φ ≡ ¬(∀(λy.¬(Φ y)))"
abbreviation fExistsBinder (binder "∃" [8]9) where "∃x. φ(x) ≡ ∃φ"
```

```
consts
  fForall :: "(i ⇒ bool) ⇒ bool"
```

Functionality Tests

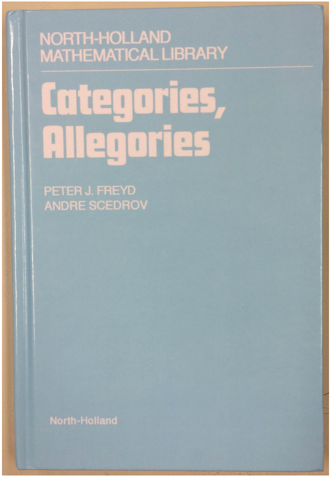


- $xrx \longrightarrow xrx$ (valid)
- $\exists y.yry \longrightarrow yry$ (countermodel)
- $(xrx \longrightarrow xrx) \longrightarrow (\exists y.yry \longrightarrow yry)$ (countermodel)
- $(xrx \longrightarrow xrx) \wedge (\exists y.y = y) \longrightarrow (\exists y.yry \longrightarrow yry)$ (valid)
- … see paper …



Part C: Application of Free Logic: Categories and Allegories (Freyd and Scedrov, 1990)

Application: Cats & Alligators



1.1. BASIC DEFINITIONS

The theory of CATEGORIES is given by two unary operations and a binary partial operation. In most contexts lower-case variables are used for the 'individuals' which are called *morphisms* or *maps*. The values of the operations are denoted and pronounced as:

$\Box x$ *the source of* x ,

$x\Box$ *the target of* x ,

xy *the composition of* x *and* y .

The axioms:

A1 xy *is defined* *iff* $x\Box = \Box y$,

A2a $(\Box x)\Box = \Box x$ *and* $\Box(x\Box) = x\Box$, A2b

A3a $(\Box x)x = x$ *and* $x(x\Box) = x$, A3b

A4a $\Box(xy) = \Box(x(\Box y))$ *and* $(xy)\Box = ((x\Box)y)\Box$, A4b

A5 $x(yz) = (xy)z$.

1.11. The ordinary equality sign $=$ will be used only in the symmetric sense, to wit: if either side is defined then so is the other and they are equal. A theory, such as this, built on an ordered list of partial operations, the domain of definition of each given by equations in the previous, and with all other axioms equational, is called an ESSENTIALLY ALGEBRAIC THEORY.

1.12. We shall use a venturi-tube $\succeq$ for *directed equality* which means: if the left side is defined then so is the right and they are equal. The axiom that $\Box(xy) = \Box(x(\Box y))$ is equivalent, in the presence of the earlier axioms, with $\Box(xy) \succeq \Box x$ as can be seen below.

1.13. $\Box(\Box x) = \Box x$ because $\Box(\Box x) = \Box((\Box x)\Box) = (\Box x)\Box = \Box x$. Similarly $(x\Box)\Box = x\Box$.

Application: Cats & Alligators

1.11. The ordinary equality sign = will be used only in the symmetric sense, to wit: if either side is defined then so is the other and they are equal. A theory, such as this, built on an ordered list of partial operations, the domain of definition of each given by equations in the previous, and with all other axioms equational, is called an ESSENTIALLY ALGEBRAIC THEORY.

- First (incorrect) formalisation: $x \approx y \quad := \quad (E(x) \longleftrightarrow E(y)) \wedge x = y$
- Freyd/Scedrov want Kleene-Eq: $x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$

1st Series of Experiments

Parameters: $\approx$, at least one undefined map ($\star$), definite description
Results A: Redundancies detected (in paper)
Results B: Inconsistency detected (after paper writing)

2nd Series of Experiments

Parameters: $\simeq$, no $\star$ (E maybe empty), no definite description
Results C: Consistent if all maps are defined (after paper writing)
Results D: Inconsistent if undefined map(s) exist (after paper writing)

3rd Series of Experiments

Parameters: Scott's own axioms (**Identity and Existence in Intuitionistic Logic**, 1977/79), uses non-reflexive equality in (A1) and Kleene Equality elsewhere, no $\star$ (E maybe empty), no definite description
Results E: Consistent if all maps are defined (after paper writing)
Results F: Still consistent if undefined map(s) exist (after paper writing)

Application: Cats & Alligators

> **1.11.** The ordinary equality sign = will be used only in the symmetric sense, to wit: if either side is defined then so is the other and they are equal. A theory, such as this, built on an ordered list of partial operations, the domain of definition of each given by equations in the previous, and with all other axioms equational, is called an ESSENTIALLY ALGEBRAIC THEORY.

- First (incorrect) formalisation: $x \approx y \quad := \quad (E(x) \longleftrightarrow E(y)) \wedge x = y$
- Freyd/Scedrov want Kleene-Eq: $x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$

1st Series of Experiments

Parameters: $\approx$, at least one undefined map ($\star$), definite description
Results A: Redundancies detected (in paper)
Results B: Inconsistency detected (after paper writing)

2nd Series of Experiments

Parameters: $\simeq$, no $\star$ (E maybe empty), no definite description
Results C: Consistent if all maps are defined (after paper writing)
Results D: Inconsistent if undefined map(s) exist (after paper writing)

3rd Series of Experiments

Parameters: Scott's own axioms (**Identity and Existence in Intuitionistic Logic**, 1977/79), uses non-reflexive equality in (A1) and Kleene Equality elsewhere, no $\star$ (E maybe empty), no definite description
Results E: Consistent if all maps are defined (after paper writing)
Results F: Still consistent if undefined map(s) exist (after paper writing)

Application: Cats & Alligators

1.11. The ordinary equality sign = will be used only in the symmetric sense, to wit: if either side is defined then so is the other and they are equal. A theory, such as this, built on an ordered list of partial operations, the domain of definition of each given by equations in the previous, and with all other axioms equational, is called an ESSENTIALLY ALGEBRAIC THEORY.

- First (incorrect) formalisation: $x \approx y \quad := \quad (E(x) \longleftrightarrow E(y)) \wedge x = y$
- Freyd/Scedrov want Kleene-Eq: $x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$

1st Series of Experiments

Parameters: $\approx$, at least one undefined map ($\star$), definite description
Results A: Redundancies detected (in paper)
Results B: Inconsistency detected (after paper writing)

2nd Series of Experiments

Parameters: $\simeq$, no $\star$ (E maybe empty), no definite description
Results C: Consistent if all maps are defined (after paper writing)
Results D: Inconsistent if undefined map(s) exist (after paper writing)

3rd Series of Experiments

Parameters: Scott's own axioms (**Identity and Existence in Intuitionistic Logic**, 1977/79), uses non-reflexive equality in (A1) and Kleene Equality elsewhere, no $\star$ (E maybe empty), no definite description
Results E: Consistent if all maps are defined (after paper writing)
Results F: Still consistent if undefined map(s) exist (after paper writing)

Application: Cats & Alligators

1.11. The ordinary equality sign = will be used only in the symmetric sense, to wit: if either side is defined then so is the other and they are equal. A theory, such as this, built on an ordered list of partial operations, the domain of definition of each given by equations in the previous, and with all other axioms equational, is called an ESSENTIALLY ALGEBRAIC THEORY.

- First (incorrect) formalisation: $x \approx y \quad := \quad (E(x) \longleftrightarrow E(y)) \wedge x = y$
- Freyd/Scedrov want Kleene-Eq: $x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$

1st Series of Experiments

Parameters: $\approx$, at least one undefined map ($\star$), definite description
Results A: Redundancies detected (in paper)
Results B: Inconsistency detected (after paper writing)

2nd Series of Experiments

Parameters: $\simeq$, no $\star$ (E maybe empty), no definite description
Results C: Consistent if all maps are defined (after paper writing)
Results D: Inconsistent if undefined map(s) exist (after paper writing)

3rd Series of Experiments

Parameters: Scott's own axioms (**Identity and Existence in Intuitionistic Logic**, 1977/79), uses non-reflexive equality in (A1) and Kleene Equality elsewhere, no $\star$ (E maybe empty), no definite description
Results E: Consistent if all maps are defined (after paper writing)
Results F: Still consistent if undefined map(s) exist (after paper writing)

Application: Cats & Alligators

> **1.11.** The ordinary equality sign = will be used only in the symmetric sense, to wit: if either side is defined then so is the other and they are equal. A theory, such as this, built on an ordered list of partial operations, the domain of definition of each given by equations in the previous, and with all other axioms equational, is called an ESSENTIALLY ALGEBRAIC THEORY.

- First (incorrect) formalisation: $x \approx y \quad := \quad (E(x) \longleftrightarrow E(y)) \wedge x = y$
- Freyd/Scedrov want Kleene-Eq: $x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$

1st Series of Experiments

Parameters: $\approx$, at least one undefined map ($\star$), definite description
Results A: Redundancies detected (in paper)
Results B: Inconsistency detected (after paper writing)

2nd Series of Experiments

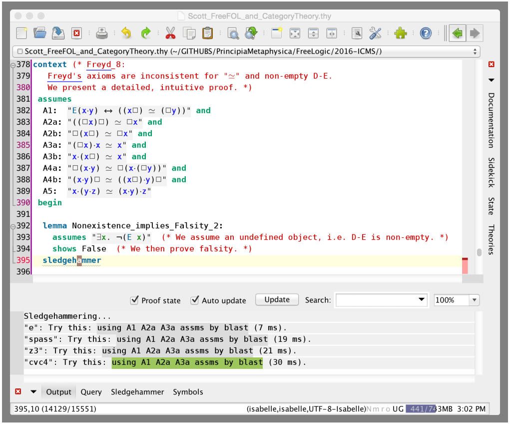
Parameters: $\simeq$, no $\star$ (E maybe empty), no definite description
Results C: Consistent if all maps are defined (after paper writing)
Results D: Inconsistent if undefined map(s) exist (after paper writing)

3rd Series of Experiments

Parameters: Scott's own axioms (**Identity and Existence in Intuitionistic Logic**, 1977/79), uses non-reflexive equality in (A1) and Kleene Equality elsewhere, no $\star$ (E maybe empty), no definite description
Results E: Consistent if all maps are defined (after paper writing)
Results F: Still consistent if undefined map(s) exist (after paper writing)

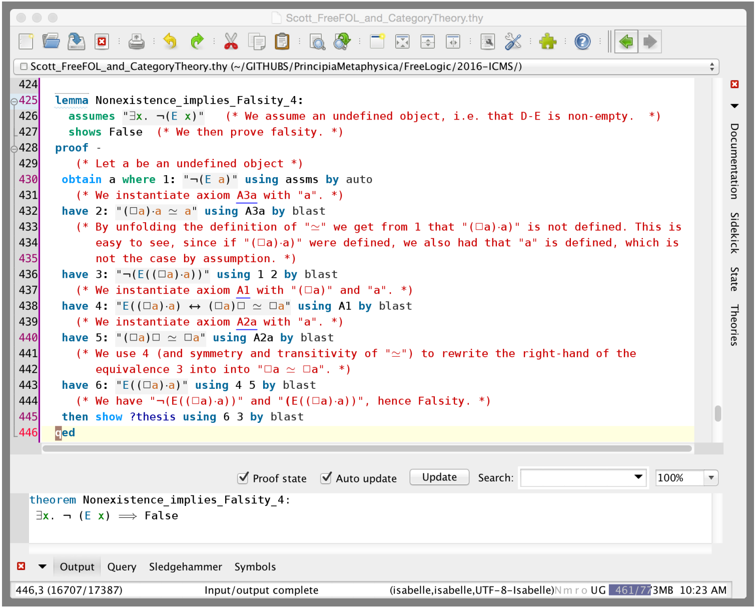
Isabelle/HOL: 2nd Series of Experiments

Parameters: $\simeq$, no $\star$ (E maybe empty), no definite description
Results C: Consistent if all maps are defined (after paper writing)
Results D: Inconsistent if undefined map(s) exist (after paper writing)



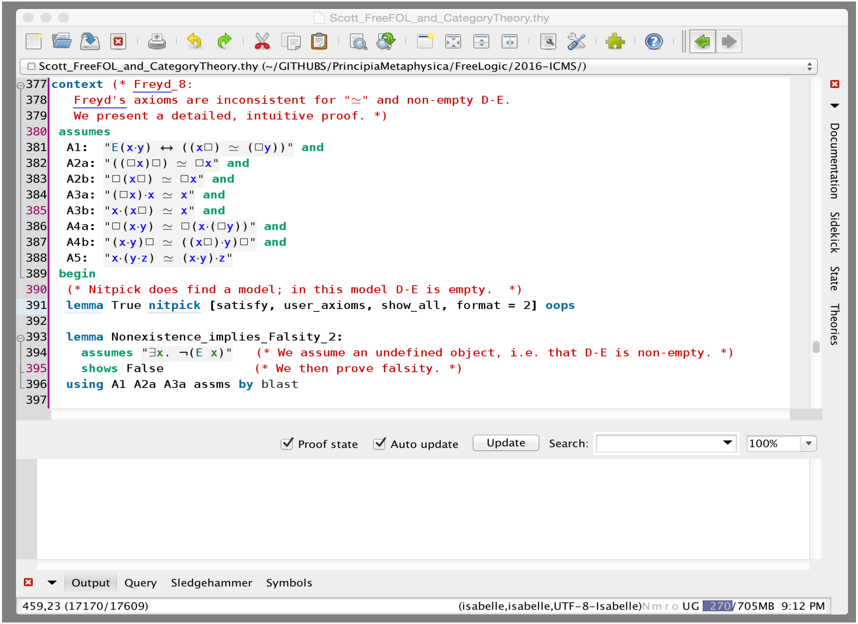
Automatic proof of “Inconsistency”.

Isabelle/HOL: 2nd Series of Experiments



Reconstruction of intuitive, interactive proof of “Inconsistency”.

Demo in Isabelle/HOL (from 2nd and 3rd series of experiments)



Reconstruction of Intuitive, Interactive Proof of "Inconsistency"

A1 xy *is defined* *iff* $x\Box = \Box y$,

A2a $(\Box x)\Box = \Box x$ *and* $\Box(x\Box) = x\Box$, A2b

A3a $(\Box x)x = x$ *and* $x(x\Box) = x$, A3b

A4a $\Box(xy) = \Box(x(\Box y))$ *and* $(xy)\Box = ((x\Box)y)\Box$, A4b

A5 $x(yz) = (xy)z$.

$$x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$$

Theorem: $(\exists x. \neg E(x)) \longrightarrow False$ **(Corollary:** $\forall x. E(x)$**)**

Proof:

Assume $\exists x. \neg E(x)$

Let a be such that $\neg E(a)$.

By instantiating $A3a$ we get $(\Box a)a \simeq a$.

By definition of $\simeq$ this means $(E((\Box a)a) \vee E(a)) \longrightarrow (E((\Box a)a) \wedge E(a) \wedge (\Box a)a = a)$.

$E((\Box a)a)$ must be false, since otherwise $E(a)$, contradicting our assumption.

Hence, $\neg E((\Box a)a)$.

By instantiating A1 with $\Box a$ and a we get $E((\Box a)a) \leftrightarrow (\Box a)\Box \simeq \Box a$.

By instantiating A2a with a we get $(\Box a)\Box \simeq \Box a$.

Hence, $E((\Box a)a)$.

Hence, Contradiction/Falsity.

Reconstruction of Intuitive, Interactive Proof of “Inconsistency”

A1 xy *is defined iff* $x\Box = \Box y$,

A2a $(\Box x)\Box = \Box x$ *and* $\Box(x\Box) = x\Box$, A2b

A3a $(\Box x)x = x$ *and* $x(x\Box) = x$, A3b

A4a $\Box(xy) = \Box(x(\Box y))$ *and* $(xy)\Box = ((x\Box)y)\Box$, A4b

A5 $x(yz) = (xy)z$.

$$x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$$

Theorem: $(\exists x.\neg E(x)) \longrightarrow False$ **(Corollary:** $\forall x.E(x)$**)**

Proof:

Assume $\exists x.\neg E(x)$

Let a be such that $\neg E(a)$.

By instantiating $A3a$ we get $(\Box a)a \simeq a$.

By definition of $\simeq$ this means $(E((\Box a)a) \vee E(a)) \longrightarrow (E((\Box a)a) \wedge E(a) \wedge (\Box a)a = a)$.

$E((\Box a)a)$ must be false, since otherwise $E(a)$, contradicting our assumption.

Hence, $\neg E((\Box a)a)$.

By instantiating A1 with $\Box a$ and a we get $E((\Box a)a) \leftrightarrow (\Box a)\Box \simeq \Box a$.

By instantiating A2a with a we get $(\Box a)\Box \simeq \Box a$.

Hence, $E((\Box a)a)$.

Hence, Contradiction/Falsity.

Reconstruction of Intuitive, Interactive Proof of “Inconsistency”

A1 xy *is defined iff* $x\Box = \Box y$,

A2a $(\Box x)\Box = \Box x$ *and* $\Box(x\Box) = x\Box$, A2b

A3a $(\Box x)x = x$ *and* $x(x\Box) = x$, A3b

A4a $\Box(xy) = \Box(x(\Box y))$ *and* $(xy)\Box = ((x\Box)y)\Box$, A4b

A5 $x(yz) = (xy)z$.

$$x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$$

Theorem: $(\exists x. \neg E(x)) \longrightarrow False$ **(Corollary:** $\forall x. E(x)$**)**

Proof:

Assume $\exists x. \neg E(x)$

Let a be such that $\neg E(a)$.

By instantiating $A3a$ we get $(\Box a)a \simeq a$.

By definition of $\simeq$ this means $(E((\Box a)a) \vee E(a)) \longrightarrow (E((\Box a)a) \wedge E(a) \wedge (\Box a)a = a)$.

$E((\Box a)a)$ must be false, since otherwise $E(a)$, contradicting our assumption.

Hence, $\neg E((\Box a)a)$.

By instantiating A1 with $\Box a$ and a we get $E((\Box a)a) \leftrightarrow (\Box a)\Box \simeq \Box a$.

By instantiating A2a with a we get $(\Box a)\Box \simeq \Box a$.

Hence, $E((\Box a)a)$.

Hence, Contradiction/Falsity.

Reconstruction of Intuitive, Interactive Proof of “Inconsistency”

A1 xy *is defined* *iff* $x\Box = \Box y$,

A2a $(\Box x)\Box = \Box x$ *and* $\Box(x\Box) = x\Box$, A2b

A3a $(\Box x)x = x$ *and* $x(x\Box) = x$, A3b

A4a $\Box(xy) = \Box(x(\Box y))$ *and* $(xy)\Box = ((x\Box)y)\Box$, A4b

A5 $x(yz) = (xy)z$.

$$x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$$

Theorem: $(\exists x.\neg E(x)) \longrightarrow False$ **(Corollary:** $\forall x.E(x)$**)**

Proof:

Assume $\exists x.\neg E(x)$

Let a be such that $\neg E(a)$.

By instantiating $A3a$ we get $(\Box a)a \simeq a$.

By definition of $\simeq$ this means $(E((\Box a)a) \vee E(a)) \longrightarrow (E((\Box a)a) \wedge E(a) \wedge (\Box a)a = a)$.

$E((\Box a)a)$ must be false, since otherwise $E(a)$, contradicting our assumption.

Hence, $\neg E((\Box a)a)$.

By instantiating A1 with $\Box a$ and a we get $E((\Box a)a) \leftrightarrow (\Box a)\Box \simeq \Box a$.

By instantiating A2a with a we get $(\Box a)\Box \simeq \Box a$.

Hence, $E((\Box a)a)$.

Hence, Contradiction/Falsity.

3rd Series of Experiments: Scott's Alternative Axioms are Consistent

Freyd/Scedrov axioms

A1 xy *is defined* *iff* $x\Box = \Box y$,

A2a $(\Box x)\Box = \Box x$ *and* $\Box(x\Box) = x\Box$, A2b

A3a $(\Box x)x = x$ *and* $x(x\Box) = x$, A3b

A4a $\Box(xy) = \Box(x(\Box y))$ *and* $(xy)\Box = ((x\Box)y)\Box$, A4b

A5 $x(yz) = (xy)z$.

$$x \simeq y \quad := \quad (E(x) \vee E(y)) \longrightarrow (E(x) \wedge E(y) \wedge x = y)$$

Scott's axioms (1977, in Freyd/Scedrov notation)

S1: $E(\Box x) \longrightarrow E(x)$ (↔ not needed)

S2: $E(x\Box) \longrightarrow E(x)$ (↔ not needed)

S3: $E(xy) \leftrightarrow x\Box =^1 \Box y$

S4: $x(yz) =^2 (xy)z$

S5: $(\Box x)x =^2 x$

S6: $x(x\Box) =^2 x$

where

$x =^1 y \quad := \quad E(x) \wedge E(y) \wedge x = y$ (non-reflexive on D)

$x =^2 y \quad := \quad (E(x) \vee E(y)) \longrightarrow x =^1 y$ ($=^2$ is $\simeq$)

Test: $(\exists x. \neg E(x)) \longrightarrow False$

Countermodel (by Nitpick).

Part D: Remark on Verified Publications

Problem: Verified Documents vs. Current Publication Practice

- Submitted document was a verified document:
  PDF was generated directly from Isabelle/HOL sources
- Springer's (re-)production process may completely mess up such papers!
- Example from a recent paper to appear in Journal of Philosophical Logic:

Problem: Verified Documents vs. Current Publication Practice

- Submitted document was a verified document:
  PDF was generated directly from Isabelle/HOL sources
- Springer's (re-)production process may completely mess up such papers!
- Example from a recent paper to appear in Journal of Philosophical Logic:

Definition 1 (Sequent calculus $\mathcal{G}_{\beta\mathfrak{fb}}$) Let Δ and Δ' be finite sets of β-normal closed formulas of HOL and let Δ, s denote the set $\Delta \cup \{s\}$. The sequent calculus $\mathcal{G}_{\beta\mathfrak{fb}}$ comprises the following rules:

Basic Rules

$$\frac{\Delta, s}{\Delta, \neg\neg s}\,\mathcal{G}(\neg) \qquad \frac{\Delta, \neg s \quad \Delta, \neg t}{\Delta, \neg(s \vee t)}\,\mathcal{G}(\vee_-) \qquad \frac{\Delta, s, t}{\Delta, (s \vee t)}\,\mathcal{G}(\vee_+)$$

$$\frac{\Delta, \neg\, (sl)\big\downarrow_\beta \quad l_\alpha \text{ closed term}}{\Delta, \neg \Pi^\alpha s}\,\mathcal{G}(\Pi_-^l) \qquad \frac{\Delta, (sc)\big\downarrow_\beta \quad c_\delta \text{ new symbol}}{\Delta, \Pi^\alpha s}\,\mathcal{G}(\Pi_+^c)$$

Initialization

$$\frac{s \text{ atomic (and } \beta\text{-normal)}}{\Delta, s, \neg s}\,\mathcal{G}(init)$$

$$\frac{\Delta, (s \doteq^o t) \quad s,t \text{ atomic}}{\Delta, \neg s, t}\,\mathcal{G}(Init^{\doteq})$$

Extensionality

$$\frac{\Delta, (\forall X_\alpha sX \doteq^\beta tX)\big\downarrow_\beta}{\Delta, (s \doteq^{\alpha\to\beta} t)}\,\mathcal{G}(\mathfrak{f}) \qquad \frac{\Delta, \neg s, t \quad \Delta, \neg t, s}{\Delta, (s \doteq^o t)}\,\mathcal{G}(\mathfrak{b})$$

Decomposition

$$\frac{\Delta, (s^1 \doteq^{\alpha_1} t^1) \ \cdots \ \Delta, (s^n \doteq^{\alpha_n} t^n) \quad \begin{array}{c} n \geq 1, \beta \in \{o, \iota\}, \\ h_{\overline{\alpha^n}\to\beta} \in \Sigma \end{array}}{\Delta, (h\overline{s^n} \doteq^\beta h\overline{t^n})}\,\mathcal{G}(d)$$

Problem: Verified Documents vs. Current Publication Practice

- Submitted document was a verified document:
  PDF was generated directly from Isabelle/HOL sources
- Springer's (re-)production process may completely mess up such papers!
- Example from a recent paper to appear in Journal of Philosophical Logic:

Definition 1 (Sequent calculus $\mathcal{G}_{\mathfrak{fb}}$) Let Δ and Δ' be finite sets of β-normal closed formulas of HOL and let Δ, s denote the set $\Delta \cup \{s\}$. The sequent calculus $\mathcal{G}_{\mathfrak{fb}}$ comprises the following rules:

Basic Rules

$$\frac{\Delta, s}{\Delta, \neg\neg s}\mathcal{G}(\neg) \quad \frac{\Delta, \neg s}{\Delta, \neg t}\Delta, \neg(s \vee t)\mathcal{G}(\vee_-) \quad \frac{\Delta, s, t}{\Delta, (s \vee t)}\mathcal{G}(\vee_+)$$

$$\frac{\Delta, \neg(sl){\downarrow_\beta} l_\alpha \text{ closed term}}{\Delta, \neg\Pi^\alpha s}(\mathcal{G}\Pi^l_-) \quad \frac{\Delta, (sc){\downarrow_\beta} \quad c_\delta \quad \text{new symbol}}{\Delta, \Pi^\alpha s}(\mathcal{G}\Pi^c_+)$$

Initialization

$$\frac{s\text{atomic}(and\beta\text{-normal})}{\Delta, s, \neg s}\mathcal{G}(init)$$

$$\frac{\Delta, (s \doteq^o t)s, t \text{ atomic}}{\Delta, \neg s, t}\mathcal{G}(Init^{\doteq})$$

Extensionality

$$\frac{\Delta, (\forall X_\alpha sX \doteq^\beta tX){\downarrow_\beta}}{\Delta, (s \doteq^{\alpha\to\beta} t)}\mathcal{G}(\mathfrak{f}) \quad \frac{\Delta, \neg s, t \Delta, \neg t, s}{\Delta, (s \doteq^o t)}\mathcal{G}(\mathfrak{b})$$

Decomposition

$$\frac{\Delta, (s^1 \doteq^{\alpha_1} t^1) \cdots \Delta, (s^n \doteq^{\alpha_n} t^n) \quad n \geq 1, \beta \in \{o, \iota\}, \; h_{\overline{\alpha^n}\to\beta} \in \Sigma}{\Delta, (h\overline{s^n} \doteq^\beta h\overline{t^n})}\mathcal{G}(d)$$

Conclusion

Achieved:

- Elegant Embedding of Free Logic in HOL
- Formalisation in Isabelle/HOL
- Very effective automation with state of the art ATPs
- Application of this framework in Category Theory
- Novel “Inconsistency” result — this time in Maths (as opposed to Philosophy)
- Scott (1977) avoids the “Inconsistency"-trap
  (once again – see Ontological Argument)

Maths can strongly benefit from rigorous formalisation!

Further Work:

- Further formalise Cats & Alligators with Scott’s axioms
- Focus on proof automation
- Analogous application/experiments in Projective Geometry
- Library of formalised Maths based on Free Logic in HOL?