Universal Logic Theorem Proving via Semantical Embeddings in HOL

Christoph Benzmüller
University of Luxemburg | Freie Universität Berlin | Saarland University

Presentation Structure

A Logic Zoo

B Universal Logic Reasoning in HOL

C Recent Success Stories

D Tutorial: Wise Men Puzzle

E Tutorial: Standard Deontic Logic

F Conclusion

You may want to download:

`https://www.cl.cam.ac.uk/research/hvg/Isabelle/`
`http://christoph-benzmueller.de/papers/2017-Calabria-Tutorial.zip`

Part A
Logic Zoo

"If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."

(Leibniz, 1677)

(A) Logic Zoo

Classical Logic, of order

0. Propositional Logic

1. First-order Logic

2. Second-order Logic

. . .

n. Higher-order Logic

Non-Classical Logics

- Intuitionistic/Constructive Logics (incl. Univalent Foundations)
- Modal Logics, Conditional Logics, Temporal Logis, Spatial Logics
- Many-valued Logics
- Paraconsistent Logics
- Free Logics, Inclusive Logics
- Logics for special applications: Ethics, Social Choice, Legal Reasoning, . . .
- . . .

(A) Logic Zoo

Classical Logic, of order

0. Propositional Logic
1. First-order Logic
2. Second-order Logic

...

n. Higher-order Logic

Non-Classical Logics

- Intuitionistic/Constructive Logics (incl. Univalent Foundations)
- Modal Logics, Conditional Logics, Temporal Logis, Spatial Logics
- Many-valued Logics
- Paraconsistent Logics
- Free Logics, Inclusive Logics
- Logics for special applications: Ethics, Social Choice, Legal Reasoning, ...
- ...

Own Research: Utilise Higher-order Logic (HOL) as Meta-Logic

This turns

- HOL into a (quite) universal logic
- HOL provers into (quite) universal reasoners

Applications in Maths, CS, AI, Philosophy, Computational Linguistics, ...

(A) Logic Zoo

Classical Logic, of order

0. Propositional Logic
1. First-order Logic
2. Second-order Logic

...

n. $\boxed{\text{Higher-order Logic}}$

Non-Classical Logics

- Intuitionistic/Constructive Logics (incl. Univalent Foundations)
- Modal Logics, Conditional Logics, Temporal Logis, Spatial Logics
- Many-valued Logics
- Paraconsistent Logics
- Free Logics, Inclusive Logics
- Logics for special applications: Ethics, Social Choice, Legal Reasoning, ...
- ...

Higher-order Logic: Don't be afraid of it!

Paradoxes (e.g. Russel's Paradox) eliminated by **Types**
Incompleteness avoided by **Henkin Semantics**

(A) Logic Zoo

Classical Logic, of order

0. Propositional Logic
1. First-order Logic
2. Second-order Logic

...

n. Higher-order Logic

Non-Classical Logics

- Intuitionistic/Constructive Logics (incl. Univalent Foundations)
- Modal Logics, Conditional Logics, Temporal Logis, Spatial Logics
- Many-valued Logics
- Paraconsistent Logics
- Free Logics, Inclusive Logics
- Logics for special applications: Ethics, Social Choice, Legal Reasoning, ...
- ...

Example Application in Metaphysics/Philosophy:

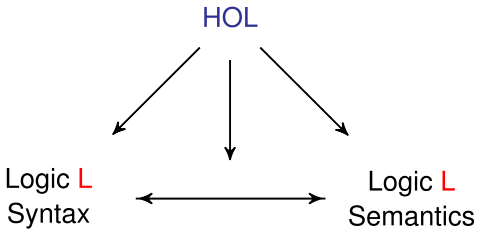
Necessarily, God exists: $\Box\exists x.Gx$

Kurt Gödel's definition of God: $Gx \;:=\; \forall\Phi.Positive\;\Phi \rightarrow \Phi x$

Part B
Universal Logic Reasoning in HOL via Shallow Semantical Embeddings

jww:
Larry Paulson (Cambridge, UK), Bruno Woltzenlogel-Paleo (ANU, Australia),
Alex Steen and Max Wisniewski (both FU Berlin)

(B) Universal Logic Reasoning in HOL



Examples for L we have already studied:

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

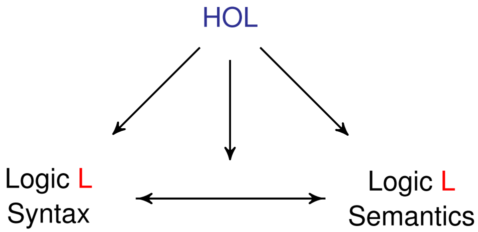
Embedding works also for quantifiers (first-order & higher-order)

HOL provers become universal logic reasoning engines!

interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, . . .

automated: TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, . . .

(B) Universal Logic Reasoning in HOL



Examples for L we have already studied:

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

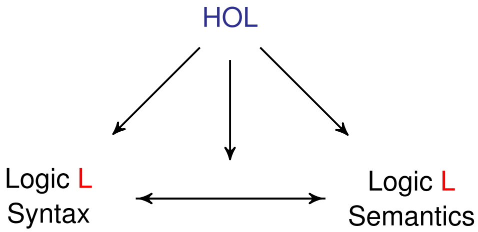
Embedding works also for quantifiers (first-order & higher-order)

HOL provers become universal logic reasoning engines!

interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, . . .

automated: TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, . . .

(B) Universal Logic Reasoning in HOL



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

Embedding works also for quantifiers (first-order & higher-order)

HOL provers become universal logic reasoning engines!

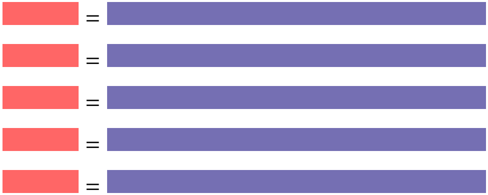
interactive:	Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, . . .
automated:	TPS, LEO-II, Satallax, Nitpick, Isabelle/HOL, . . .

(B) Universal Logic Reasoning in HOL

HOL (meta-logic) $\varphi ::=$

L (object-logic) $\psi ::=$

Embedding of in



Pass this set of equations to a HOL theorem prover

(B) Universal Logic Reasoning in HOL

HOL $s,t \quad ::= \quad c_\alpha \mid x_\alpha \mid (\lambda x_\alpha s_\beta)_{\alpha\to\beta} \mid (s_{\alpha\to\beta}\, t_\alpha)_\beta \mid \neg s_o \mid s_o \vee t_o \mid \forall x_\alpha\, t_o$

HOML $\varphi,\psi \quad ::= \quad \ldots \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \to \psi \mid \Box\varphi \mid \Diamond\varphi \mid \forall x_\gamma\, \varphi \mid \exists x_\gamma\, \varphi$

HOML in HOL: HOML formulas φ are mapped to HOL predicates $\varphi_{\mu\to o}$ (explicit representation of labelled formulas)

$$
\begin{aligned}
\neg &= \lambda\varphi_{\mu\to o}\lambda w_\mu \neg\varphi w \\
\wedge &= \lambda\varphi_{\mu\to o}\lambda\psi_{\mu\to o}\lambda w_\mu(\varphi w \wedge \psi w) \\
\to &= \lambda\varphi_{\mu\to o}\lambda\psi_{\mu\to o}\lambda w_\mu(\neg\varphi w \vee \psi w) \\
\forall &= \lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu \forall d_\gamma\, hdw \\
\exists &= \lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu \exists d_\gamma\, hdw \\
\Box &= \lambda\varphi_{\mu\to o}\lambda w_\mu \forall u_\mu\, (\neg \boldsymbol{r} wu \vee \varphi u) \\
\Diamond &= \lambda\varphi_{\mu\to o}\lambda w_\mu \exists u_\mu\, (\boldsymbol{r} wu \wedge \varphi u) \\
\text{valid} &= \lambda\varphi_{\mu\to o}\forall w_\mu \varphi w
\end{aligned}
$$

Ax (polymorphic over γ)

The equations in Ax are given as axioms to the HOL provers!

(B) Universal Logic Reasoning in HOL

HOL $s, t \quad ::= \quad \boldsymbol{c}_\alpha \mid x_\alpha \mid (\lambda x_\alpha s_\beta)_{\alpha\to\beta} \mid (s_{\alpha\to\beta}\, t_\alpha)_\beta \mid \neg s_o \mid s_o \vee t_o \mid \forall x_\alpha\, t_o$

HOML $\varphi, \psi \quad ::= \quad \ldots \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \to \psi \mid \Box\varphi \mid \Diamond\varphi \mid \forall x_\gamma\, \varphi \mid \exists x_\gamma\, \varphi$

HOML in HOL: HOML formulas φ are mapped to HOL predicates $\varphi_{\mu\to o}$
(explicit representation of labelled formulas)

$\neg$	=	$\lambda\varphi_{\mu\to o}\lambda w_\mu \neg\varphi w$
$\wedge$	=	$\lambda\varphi_{\mu\to o}\lambda\psi_{\mu\to o}\lambda w_\mu(\varphi w \wedge \psi w)$
$\to$	=	$\lambda\varphi_{\mu\to o}\lambda\psi_{\mu\to o}\lambda w_\mu(\neg\varphi w \vee \psi w)$
$\forall$	=	$\lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu \forall d_\gamma\, hdw$
$\exists$	=	$\lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu \exists d_\gamma\, hdw$
$\Box$	=	$\lambda\varphi_{\mu\to o}\lambda w_\mu \forall u_\mu\, (\neg \boldsymbol{r}wu \vee \varphi u)$
$\Diamond$	=	$\lambda\varphi_{\mu\to o}\lambda w_\mu \exists u_\mu\, (\boldsymbol{r}wu \wedge \varphi u)$
valid	=	$\lambda\varphi_{\mu\to o}\forall w_\mu \varphi w$

Ax (polymorphic over γ)

The equations in Ax are given as axioms to the HOL provers!

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond \exists x G x$
HOML formula embedded in HOL	valid $(\Diamond \exists x G x)$
expansion	$(\lambda \varphi \forall w_\mu \varphi\, w)(\Diamond \exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu ((\Diamond \exists x G x)\, w)$
expansion	$\forall w_\mu (((\lambda \varphi \lambda w_\mu \exists u_\mu\, (rwu \wedge \varphi u)) \exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (rwu \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu (rwu \wedge (\exists (\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu (rwu \wedge ((\lambda h_{\gamma \to (\mu \to o)} \lambda w_\mu \exists d_\gamma\, hdw)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (rwu \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond\exists x G x$
HOML formula embedded in HOL	valid $(\Diamond\exists x G x)$
expansion	$(\lambda\varphi\forall w_\mu \varphi\, w)(\Diamond\exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu((\Diamond\exists x G x)\, w)$
expansion	$\forall w_\mu(((\lambda\varphi\lambda w_\mu \exists u_\mu\, (rwu \wedge \varphi u))\exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu(rwu \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu(rwu \wedge (\exists(\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu(rwu \wedge ((\lambda h_{\gamma\to(\mu\to o)} \lambda w_\mu \exists d_\gamma\, hdw)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu(rwu \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond \exists x G x$
HOML formula embedded in HOL	valid $(\Diamond \exists x G x)$
expansion	$(\lambda \varphi \forall w_\mu \varphi\, w)(\Diamond \exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu ((\Diamond \exists x G x)\, w)$
expansion	$\forall w_\mu (((\lambda \varphi \lambda w_\mu \exists u_\mu\, (rwu \wedge \varphi u)) \exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (rwu \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu (rwu \wedge (\exists (\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu (rwu \wedge ((\lambda h_{\gamma \to (\mu \to o)} \lambda w_\mu \exists d_\gamma\, hdw)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (rwu \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond\exists x G x$
HOML formula embedded in HOL	valid $(\Diamond\exists x G x)$
expansion	$(\lambda\varphi\forall w_\mu \varphi\, w)(\Diamond\exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu((\Diamond\exists x G x)\, w)$
expansion	$\forall w_\mu(((\lambda\varphi\lambda w_\mu \exists u_\mu\, (rwu \wedge \varphi u))\exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (rwu \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu (rwu \wedge (\exists(\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu (rwu \wedge ((\lambda h_{\gamma\to(\mu\to o)} \lambda w_\mu \exists d_\gamma\, hdw)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (rwu \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond \exists x G x$
HOML formula embedded in HOL	valid $(\Diamond \exists x G x)$
expansion	$(\lambda \varphi \forall w_\mu \varphi\, w)(\Diamond \exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu ((\Diamond \exists x G x)\, w)$
expansion	$\forall w_\mu (((\lambda \varphi \lambda w_\mu \exists u_\mu\, (\boldsymbol{r} w u \wedge \varphi u)) \exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge (\exists (\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge ((\lambda h_{\gamma \to (\mu \to o)} \lambda w_\mu \exists d_\gamma\, h d w)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond\exists x G x$
HOML formula embedded in HOL	valid $(\Diamond\exists x G x)$
expansion	$(\lambda\varphi\forall w_\mu \varphi\, w)(\Diamond\exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu((\Diamond\exists x G x)\, w)$
expansion	$\forall w_\mu(((\lambda\varphi\lambda w_\mu\exists u_\mu\, (\boldsymbol{r} w u \wedge \varphi u))\exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu\exists u_\mu(\boldsymbol{r} w u \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu\exists u_\mu(\boldsymbol{r} w u \wedge (\exists(\lambda x G x)) u)$
expansion	$\forall w_\mu\exists u_\mu(\boldsymbol{r} w u \wedge ((\lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu\exists d_\gamma\, h d w)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu\exists u_\mu(\boldsymbol{r} w u \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond \exists x G x$
HOML formula embedded in HOL	valid $(\Diamond \exists x G x)$
expansion	$(\lambda\varphi \forall w_\mu \varphi\, w)(\Diamond \exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu ((\Diamond \exists x G x)\, w)$
expansion	$\forall w_\mu (((\lambda\varphi \lambda w_\mu \exists u_\mu\, (\boldsymbol{r} w u \wedge \varphi u)) \exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge (\exists (\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge ((\lambda h_{\gamma\to(\mu\to o)} \lambda w_\mu \exists d_\gamma\, h d w)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond\exists x G x$
HOML formula embedded in HOL	valid $(\Diamond\exists x G x)$
expansion	$(\lambda\varphi\forall w_\mu \varphi\, w)(\Diamond\exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu((\Diamond\exists x G x)\, w)$
expansion	$\forall w_\mu(((\lambda\varphi\lambda w_\mu \exists u_\mu\, (\boldsymbol{r} wu \wedge \varphi u))\exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu(\boldsymbol{r} wu \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu(\boldsymbol{r} wu \wedge (\exists(\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu(\boldsymbol{r} wu \wedge ((\lambda h_{\gamma\to(\mu\to o)} \lambda w_\mu \exists d_\gamma\, hdw)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu(\boldsymbol{r} wu \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond \exists x G x$
HOML formula embedded in HOL	valid $(\Diamond \exists x G x)$
expansion	$(\lambda \varphi \forall w_\mu \varphi\, w)(\Diamond \exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu ((\Diamond \exists x G x)\, w)$
expansion	$\forall w_\mu (((\lambda \varphi \lambda w_\mu \exists u_\mu\, (\boldsymbol{r} w u \wedge \varphi u)) \exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge (\exists (\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge ((\lambda h_{\gamma \to (\mu \to o)} \lambda w_\mu \exists d_\gamma\, h d w)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu (\boldsymbol{r} w u \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

(B) Universal Logic Reasoning in HOL

Example

HOML formula	$\Diamond\exists x G x$
HOML formula embedded in HOL	valid $(\Diamond\exists x G x)$
expansion	$(\lambda\varphi \forall w_\mu \varphi\, w)(\Diamond\exists x G x)$
$\beta\eta$-normalisation	$\forall w_\mu((\Diamond\exists x G x)\, w)$
expansion	$\forall w_\mu(((\lambda\varphi \lambda w_\mu \exists u_\mu\, (\boldsymbol{r} w u \wedge \varphi u))\exists x G x)\, w)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu(\boldsymbol{r} w u \wedge (\exists x G x) u)$
syntactic sugar	$\forall w_\mu \exists u_\mu(\boldsymbol{r} w u \wedge (\exists(\lambda x G x)) u)$
expansion	$\forall w_\mu \exists u_\mu(\boldsymbol{r} w u \wedge ((\lambda h_{\gamma\rightarrow(\mu\rightarrow o)} \lambda w_\mu \exists d_\gamma\, h d w)(\lambda x G x)) u)$
$\beta\eta$-normalisation	$\forall w_\mu \exists u_\mu(\boldsymbol{r} w u \wedge \exists x G x u)$

What are we doing?

In order to prove that φ is valid in HOML,
–> we instead prove that valid φ can be derived from Ax in HOL.

This can be done with interactive or automated HOL theorem provers.
(Isabelle, HOL4, HOL-light, Coq, LEO-II, Leo-III, Lean, etc.)

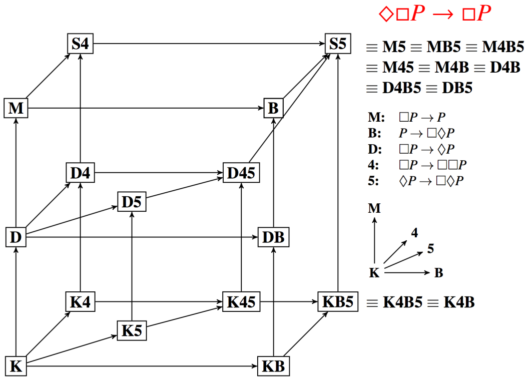
(B) Universal Logic Reasoning in Isabelle/HOL

GodProof.thy (~/chris/trunk/tex/talks/2016-BMS/)

```
theory GodProof imports Main
begin
 typedecl i -- "type for possible worlds"
 typedecl μ -- "type for individuals"
 type_synonym σ = "(i⇒bool)"

(* Shallow embedding modal logic connectives in HOL *)
 abbreviation mneg ("¬_"[52]53)     where "¬φ    ≡ λw. ¬φ(w)"
 abbreviation mand (infixr"∧"51)    where "φ∧ψ   ≡ λw. φ(w)∧ψ(w)"
 abbreviation mor (infixr"∨"50)     where "φ∨ψ   ≡ λw. φ(w)∨ψ(w)"
 abbreviation mimp (infixr"→"49)    where "φ→ψ ≡ λw. φ(w)⟶ψ(w)"
 abbreviation mequ (infixr"↔"48)    where "φ↔ψ ≡ λw. φ(w)⟷ψ(w)"
 abbreviation mnegpred ("¬_"[52]53) where "¬Φ   ≡ λx.λw. ¬Φ(x)(w)"

(* Shallow embedding of generic box and diamnond operators *)
 abbreviation mboxgen ("□") where "□r φ ≡ λw. ∀v.  r w v ⟶ φ(v)"
 abbreviation mdiagen ("◇") where "◇r φ ≡ λw. ∃v. r w v ∧ φ(v)"

(* Shallow embedding of constant domain quantifiers in HOL *)
 abbreviation mall_const ("∀c") where "∀c Φ ≡ λw.∀x. Φ(x)(w)"
 abbreviation mallB_const (binder"∀c"[8]9) where "∀c x. φ(x) ≡ ∀c φ"
 abbreviation mexi_const ("∃c") where "∃c Φ ≡ λw.∃x. Φ(x)(w)"
 abbreviation mexiB_const (binder"∃c"[8]9) where "∃c x. φ(x) ≡ ∃c φ"
```

(B) Universal Logic Reasoning in HOL

Which modal logic?



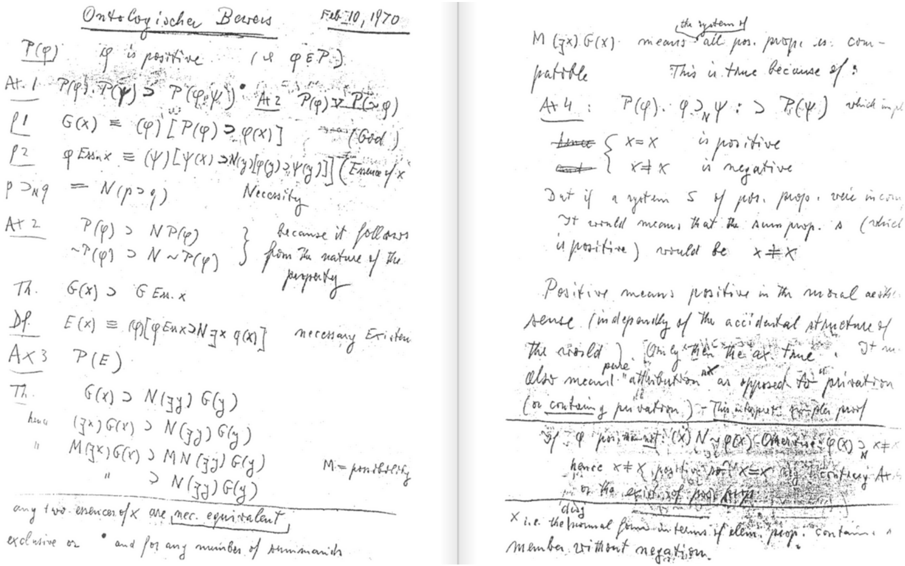
Which notion of quantification?

- possibilist quantifiers — constant domain semantics
- actualist quantifiers — varying domain semantics

Computational Metaphysics: See our Recent Publications

- **Computer-Assisted Analysis of the Anderson-Hájek Controversy**, In Logica Universalis, 2017.
- **Analysis of an Ontological Proof Proposed by Leibniz**, Chapter in Death and Anti-Death, Volume 14, Ria University Press, 2017.
- **An Object-Logic Explanation for the Inconsistency in Gödel's Ontological Theory**, KI 2016, Springer, LNCS, 2016.
- **The Inconsistency in Gödel's Ontological Argument: A Success Story for AI in Metaphysics**, IJCAI 2016, AAAI Press, 2016.
- **The Modal Collapse as a Collapse of the Modal Square of Opposition**, Chapter in The Square of Opposition: A Cornerstone of Thought (Collection of papers related to the World Congress on the Square of Opposition IV, Vatican, 2014), Springer, Studies in Universal Logic, 2016.
- **On Logic Embeddings and Gödel's God**, WADT 2014, Springer, LNCS, 2015.
- **Experiments in Computational Metaphysics: Gödel's Proof of God's Existence**, In Science & Spiritual Quest, 9th All India Students' Conference, Bhaktivedanta Institute, Kolkata, 2015.
- **Invited Talk: On a (Quite) Universal Theorem Proving Approach and Its Application in Metaphysics**, TABLEAUX 2015, Springer, LNAI, 2015.
- **Automating Gödel's Ontological Proof of God's Existence with Higher-order Automated Theorem Provers**, ECAI 2014, IOS Press, Frontiers in Artificial Intelligence and Applications, 2014.
- ...
- Further papers in preparation

Computational Metaphysics: Kurt Gödel's Ontological Argument



Ontologischer Beweis   Feb 10, 1970

$P(\varphi)$ φ is positive   (or $\varphi \in P$)

At. 1 $P(\varphi).P(\psi) \supset P(\varphi.\psi)$ •   At 2 $P(\varphi) \veebar P(\sim\varphi)$

P1 $G(x) \equiv (\varphi)[P(\varphi) \supset \varphi(x)]$ (God)

P2 $\varphi \text{ Ess.} x \equiv (\psi)[\psi(x) \supset N(y)[\varphi(y) \supset \psi(y)]]$ (Essence of x)

$p \supset_N q = N(p \supset q)$ Necessity

At 2 $P(\varphi) \supset NP(\varphi)$, $\sim P(\varphi) \supset N\sim P(\varphi)$ } because it follows from the nature of the property

Th. $G(x) \supset G \text{ Ess.} x$

Df. $E(x) \equiv (\varphi)[\varphi \text{ Ess } x \supset N \exists x\, \varphi(x)]$ necessary Existence

Ax 3 $P(E)$

Th. $G(x) \supset N(\exists y)\, G(y)$

hence $(\exists x) G(x) \supset N(\exists y)\, G(y)$

" $M(\exists x) G(x) \supset MN(\exists y)\, G(y)$ M = possibility

" $\supset N(\exists y)\, G(y)$

any two essences of x are nec. equivalent

exclusive or   • and for any number of summands

$M(\exists x) G(x)$ means the system of all pos. props. is compatible. This is true because of:

At 4: $P(\varphi).\ \varphi \supset_N \psi : \supset P(\psi)$ which impl.

~~true~~ { $x = x$ is positive
~~not~~ { $x \neq x$ is negative

But if a system S of pos. props. were incons. it would mean that the sum prop. s (which is positive) would be $x \neq x$

Positive means positive in the moral aesth. sense (independently of the accidental structure of the world). Only then the at. true. It may also mean pure "attribution"* as opposed to "privation" (or containing privation). This interpret. simpler proof

If φ pos. not: $(x) N \sim \varphi(x)$. Otherwise: $\varphi(x) \supset_N x \neq x$ hence $x \neq x$ positive not $x = x$ neg. contrary At. or the exist. of pos. prop.

* i.e. the disj. normal form in terms of elem. prop. contains a member without negation.

Computational Metaphysics: Scott's Variant of Gödel's Ontological Argument

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive:
$\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists xG(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists xG(x)$

Computational Metaphysics: Scott's Variant of Gödel's Ontological Argument

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists xG(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \boxed{\phi(x) \wedge} \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists xG(x)$

Difference to Gödel (who omits this conjunct)

Computational Metaphysics: Scott's Variant of Gödel's Ontological Argument

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive:
$$\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists xG(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists xG(x)$

Modal operators are used

Computational Metaphysics: Scott's Variant of Gödel's Ontological Argument

Axiom A1 Either a property or its negation is positive, but not both: $\forall \phi [P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive: $\boxed{\forall \phi \forall \psi} [(P(\phi) \wedge \Box \forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall \phi [P(\phi) \rightarrow \Diamond \exists x \phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \boxed{\forall \phi}[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond \exists x G(x)$

Axiom A4 Positive properties are necessarily positive: $\forall \phi [P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi \; ess.\; x \leftrightarrow \phi(x) \wedge \forall \psi(\psi(x) \rightarrow \Box \forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x [G(x) \rightarrow G \; ess.\; x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall \phi [\phi \; ess.\; x \rightarrow \Box \exists y \phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box \exists x G(x)$

second-order quantifiers

Computational Metaphysics: Vision of Leibniz (1646–1716) — *Calculemus!*

If controversies were to arise, there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other . . . : Let us calculate.

(Translation by Russell)

Quo facto, quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos Computistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo . . . dicere: calculemus. (Leibniz, 1684)

Required:
characteristica universalis and **calculus ratiocinator**

Computational Metaphysics: Scott's and Gödel's Variants — Demo

Axiom A1 Either a property or its negation is positive, but not both: $\forall \phi [P(\neg \phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive:
$$\forall \phi \forall \psi [(P(\phi) \wedge \Box \forall x [\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$$

Thm. T1 Positive properties are possibly exemplified: $\forall \phi [P(\phi) \rightarrow \Diamond \exists x \phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall \phi [P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond \exists x G(x)$

Axiom A4 Positive properties are necessarily positive: $\forall \phi [P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi \; ess. \; x \leftrightarrow \phi(x) \wedge \forall \psi (\psi(x) \rightarrow \Box \forall y (\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x [G(x) \rightarrow G \; ess. \; x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall \phi [\phi \; ess. \; x \rightarrow \Box \exists y \phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box \exists x G(x)$

Computational Metaphysics: Scott's and Gödel's Variants — Demo

Axiom A1 $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2

$$\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$$

Thm. T1 $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 $P(G)$

Cor. C $\Diamond\exists x G(x)$

Axiom A4 $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2

$$\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$$

Thm. T2 $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3

$$NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$$

Axiom A5 $P(NE)$

Thm. T3 $\Box\exists x G(x)$

Computational Metaphysics: Scott's and Gödel's Variants — Demo

Axiom A1 $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2

$$\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$$

Def. D1 $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 $P(G)$

Axiom A4 $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2

$$\phi\ ess.\ x \leftrightarrow \phi(x)\ \wedge\ \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$$

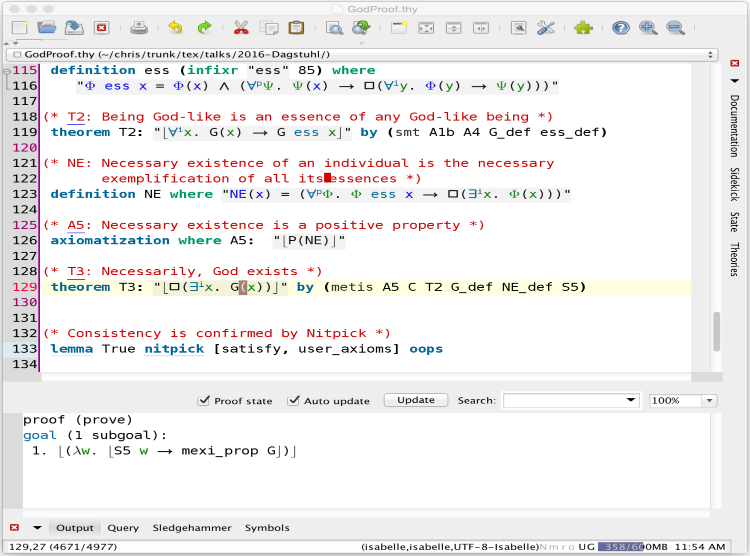
Def. D3

$$NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$$

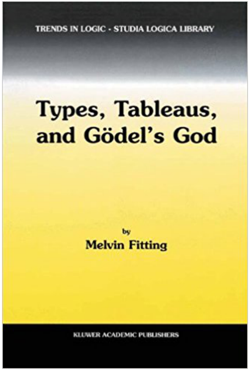
Axiom A5 $P(NE)$

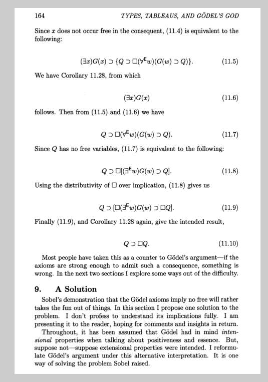
Thm. T3 $\Box\exists x G(x)$

Computational Metaphysics: Scott's and Gödel's Variants — Demo



Computational Metaphysics: Intensional Version of Ontological Argument





164 *TYPES, TABLEAUS, AND GÖDEL'S GOD*

Since x does not occur free in the consequent, (11.4) is equivalent to the following:

$$(\exists x)G(x) \supset \{Q \supset \Box(\forall^{\mathsf{E}} w)(G(w) \supset Q)\}. \tag{11.5}$$

We have Corollary 11.28, from which

$$(\exists x)G(x) \tag{11.6}$$

follows. Then from (11.5) and (11.6) we have

$$Q \supset \Box(\forall^{\mathsf{E}} w)(G(w) \supset Q). \tag{11.7}$$

Since Q has no free variables, (11.7) is equivalent to the following:

$$Q \supset \Box[(\exists^{\mathsf{E}} w)G(w) \supset Q]. \tag{11.8}$$

Using the distributivity of $\Box$ over implication, (11.8) gives us

$$Q \supset [\Box(\exists^{\mathsf{E}} w)G(w) \supset \Box Q]. \tag{11.9}$$

Finally (11.9), and Corollary 11.28 again, give the intended result,

$$Q \supset \Box Q. \tag{11.10}$$

Most people have taken this as a counter to Gödel's argument—if the axioms are strong enough to admit such a consequence, something is wrong. In the next two sections I explore some ways out of the difficulty.

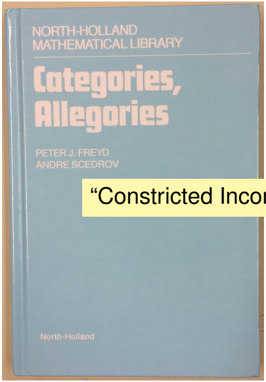
9. A Solution

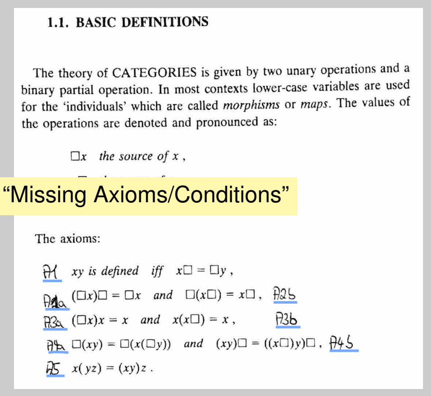
Sobel's demonstration that the Gödel axioms imply no free will rather takes the fun out of things. In this section I propose one solution to the problem. I don't profess to understand its implications fully. I am presenting it to the reader, hoping for comments and insights in return.

Throughout, it has been assumed that Gödel had in mind *intensional* properties when talking about positiveness and essence. But, suppose not—suppose extensional properties were intended. I reformulate Gödel's argument under this alternative interpretation. It is one way of solving the problem Sobel raised.

- Objective: Formalisation/Automation of Fitting's Book
- Joint work with: David Fuenmayor (student of mine)
- Logics: Intensional Higher-Order Modal Logic
- Results:
  - essential parts are formalised (by David), only minor issues detected so far
  - good degree of automation
- Sources: `https://github.com/cbenzmueller/TypesTableauxAndGoedelsGod`

Theory Exploration in Category Theory





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 ,

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

"Constricted Inconsistency" or "Missing Axioms/Conditions"

- Joint work with: Dana Scott (Berkeley) and students at FU Berlin
- Logics: Free Logic — well suited for modeling undefinedness and partiality
- **Results:**
  - development of 6 related (equivalent) axiom systems for category theory
  - from Monoids . . . via Scott's (1977) axiom system . . . to Freyd/Scedrov (1990)
  - **for Freyd and Scedrov (1990) we revealed some flaws/issues**
- Further Reading: [Benzmüller&Scott, ICMS'2016], [Benzmüller&Scott, arXiv'2016]

Computational Metaphysics: Principia Metaphysica (Zalta)

"Inconsistency/Paradox detected"

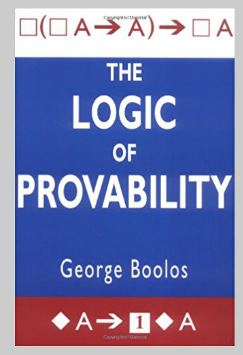
NOTE: This is an excerpt from an incomplete draft of the monograph *Principia Logica-Metaphysica*. The monograph currently has four parts:

Part I: Prophilosophy
Part II: Philosophy
Part III: Metaphilosophy
Part IV: Technical Appendices, Bibliography, Index

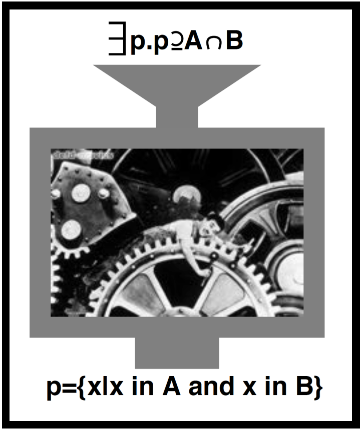
This excerpt was generated on October 18, 2016 and contains:

Principia Logico-Metaphysica

- Objective: Formalisation/Automation of Principia Metaphysica
- Joint work with: Daniel Kirchner (student of mine), Ed Zalta (Stanford)
- Logics: Hyper-Intensional Higher-Order Modal Logic
- Results:
  - PLM fully formalised (by Daniel)
  - Paradox/Inconsisteny detected
  - good degree of automation, abstract level theorem prover
- Sources: `https://github.com/ekpyron/TAO`



- Objective: Formalisation/Automation of Boolos' Book
- Joint work with: David Streit (student of mine)
- Logics: Provability Logic
- Results:
  - essential parts are already formalised, no issues detected so far
  - good degree of automation
- Sources: `will-soon-be-made-public`



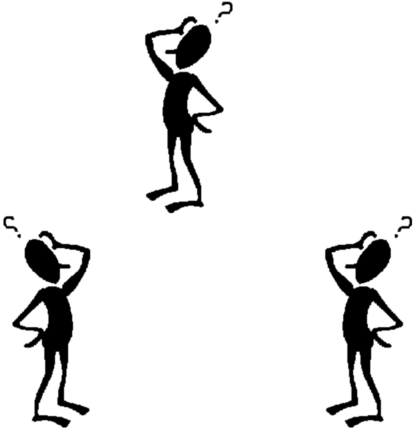
Part D
Tutorial Wise Men Puzzle

Can you represent and solve the following problem?

<u>Wise Men Puzzle</u>

Once upon a time, a king wanted to find the wisest out of his three wisest men. He arranged them in a circle and told them that he would put a white or a black spot on their foreheads and that one of the three spots would certainly be white. The three wise men could see and hear each other but, of course, they could not see their faces reflected anywhere. The king, then, asked to each of them to find out the color of his own spot. After a while, the wisest correctly answered that his spot was white.

How could he know that?



Can you represent and solve the following problem?

Wise Men Puzzle

Once upon a time, a king wanted to find the wisest out of his three wisest men. He arranged them in a circle and told them that he would put a white or a black spot on their foreheads and that one of the three spots would certainly be white. The three wise men could see and hear each other but, of course, they could not see their faces reflected anywhere. The king, then, asked to each of them to find out the color of his own spot. After a while, the wisest correctly answered that his spot was white.

How could he know that?

	A	B	C	Answer
1	*	b	b	A w
2	b	*	b	B w
3	b	b	*	C w
4	*	w	b	A w (⇒ 2.)
5	*	b	w	A w (⇒ 3.)
6	w	*	b	B w (⇒ 1.)
7	b	*	w	B w (⇒ 3.)
8	w	b	*	C w (⇒ 1.)
9	b	w	*	C w (⇒ 2.)
10	*	w	w	A w (⇒ 8.)
11	w	*	w	B w (⇒ 8.)
12	w	w	*	C w (⇒ 4.)

Can you represent and solve the following problem?

Wise Men Puzzle

Once upon a time, a king wanted to find the wisest out of his three wisest men. He arranged them in a circle and told them that he would put a white or a black spot on their foreheads and that one of the three spots would certainly be white. The three wise men could see and hear each other but, of course, they could not see their faces reflected anywhere. The king, then, asked to each of them to find out the color of his own spot. After a while, the wisest correctly answered that his spot was white.

How could he know that?

Representation in our approach

- Knowledge of wise men: KT45-operators $\Box_a, \Box_b, \Box_c$
- Mutual knowledge of a, b, c:

$$\Box_{Eabc}\varphi = \Box_a\varphi \wedge \Box_b\varphi \wedge \Box_c\varphi$$

$$Eabc = a \cup b \cup c$$

- Mutual knowledge is not a knowledge operator: reflexive (T), but not necessarily transitive (4) or euclidean (5)
- KT45-operator for common knowledge of group a, b, c can easily be introduced by taking the *transitive closure* of $Eabc$

Can you represent and solve the following problem?

Wise Men Puzzle

Once upon a time, a king wanted to find the wisest out of his three wisest men. He arranged them in a circle and told them that he would put a white or a black spot on their foreheads and that one of the three spots would certainly be white. The three wise men could see and hear each other but, of course, they could not see their faces reflected anywhere. The king, then, asked to each of them to find out the color of his own spot. After a while, the wisest correctly answered that his spot was white.

How could he know that?

Representation in our approach

- for $i = a, b, c$:

Nec	from φ follows $\Box_i\varphi$	
K	$\Box_i(\varphi \rightarrow \psi) \rightarrow (\Box_i\varphi \rightarrow \Box_i\psi)$	
T	$\Box_i\varphi \rightarrow \varphi$	'veridicality' or 'truth'
4	$\Box_i\varphi \rightarrow \Box_i\Box_i\varphi$	'positive introspection'
5	$\neg\Box_i\varphi \rightarrow \Box_i\neg\Box_i\varphi$	'negative introspection'

- Mutual knowledge:
  $Eabc = a \cup b \cup c$
- Common knowledge:
  $Cabc = transitiveClosure(Eabc)$
-

Nec	from φ follows $\Box_{Cabc}\varphi$
K	$\Box_{Cabc}(\varphi \rightarrow \psi) \rightarrow (\Box_{Cabc}\varphi \rightarrow \Box_{Cabc}\psi)$
T	$\Box_{Cabc}\varphi \rightarrow \varphi$
4	$\Box_{Cabc}\varphi \rightarrow \Box_{Cabc}\Box_{Cabc}\varphi$
5	$\neg\Box_{Cabc}\varphi \rightarrow \Box_{Cabc}\neg\Box_{Cabc}\varphi$

Can you represent and solve the following problem?

Wise Men Puzzle

Once upon a time, a king wanted to find the wisest out of his three wisest men. He arranged them in a circle and told them that he would put a white or a black spot on their foreheads and that one of the three spots would certainly be white. The three wise men could see and hear each other but, of course, they could not see their faces reflected anywhere. The king, then, asked to each of them to find out the color of his own spot. After a while, the wisest correctly answered that his spot was white.

How could he know that?

Representation in our approach

- $\Box_{Cabc}(ws\ a \vee ws\ b \vee ws\ c)$
- for $x \neq y \in \{a,b,c\}$ we have $\Box_{Cabc}(ws\ x \longrightarrow \Box_y(ws\ x))$
- for $x \neq y \in \{a,b,c\}$ we have $\Box_{Cabc}(\neg(ws\ x) \longrightarrow \Box_y\neg(ws\ x))$
- $\Box_{Cabc}\neg\Box_a(ws\ a)$
- $\Box_{Cabc}\neg\Box_b(ws\ b)$
- Show: $\Box_c(ws\ c)$

Can you represent and solve the following problem?

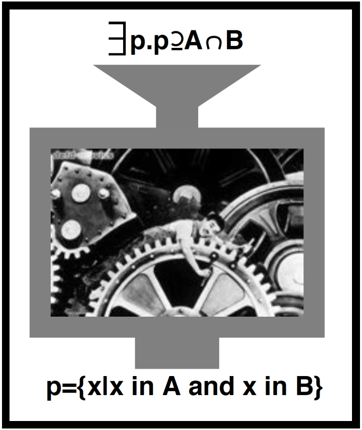
Wise Men Puzzle

Once upon a time, a king wanted to find the wisest out of his three wisest men. He arranged them in a circle and told them that he would put a white or a black spot on their foreheads and that one of the three spots would certainly be white. The three wise men could see and hear each other but, of course, they could not see their faces reflected anywhere. The king, then, asked to each of them to find out the color of his own spot. After a while, the wisest correctly answered that his spot was white.

How could he know that?

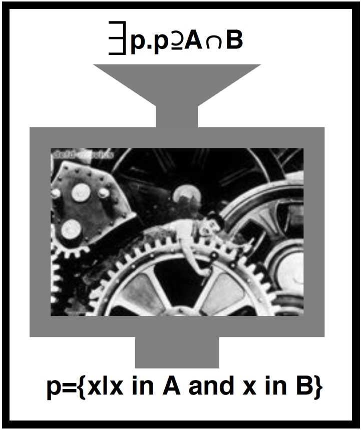
Representation in our approach

- $\Box_t \Box_{Cabc}(ws\ a \lor ws\ b \lor ws\ c)$
- for $x \neq y \in \{a, b, c\}$ we have $\Box_t \Box_{Cabc}(ws\ x \rightarrow \Box_y(ws\ x))$
- for $x \neq y \in \{a, b, c\}$ we have $\Box_t \Box_{Cabc}(\neg(ws\ x) \rightarrow \Box_y \neg(ws\ x))$
- $\Box_t \Box_{Cabc} \neg \Box_a(ws\ a)$
- $\Box_t \Box_t \Box_{Cabc} \neg \Box_b(ws\ b)$
- Show: $\Box_t \Box_t \Box_t \Box_c(ws\ c)$



Demo/Tutorial in Isabelle/HOL

http://christoph-benzmueller.de/papers/2017-Calabria-Tutorial.zip



Part E
Standard Deontic Logic
Demo/Tutorial in Isabelle/HOL

http://christoph-benzmueller.de/papers/2017-Calabria-Tutorial.zip

Standard Deontic Logic

Representation in our approach:

modal KD-operator: **OB** (it is obligatory that)

Nec	from φ follows $\mathbf{OB}\varphi$
K	$\mathbf{OB}(\varphi \rightarrow \psi) \rightarrow (\mathbf{OB}\varphi \rightarrow \mathbf{OB}\psi)$
D	$\mathbf{OB}\varphi \rightarrow \neg\mathbf{OB}\neg\varphi$ seriality

Conclusion

HOL as a (quite) Universal Metalogic via Shallow Semantic Embeddings:

- Very lean approach to integrate and combine logics
- High degree of automation (via the embeddings)
- Works surprisingly well in (our recent) practical applications
- Intuitive interaction at abstract level supported by proof assistants
- Novel results in different application domains

Approach has recently been used in my interdisciplinary lecture course on Computational Metaphysics (which won the central teaching award of FU Berlin)

Recent work in Luxemburg: Application to Normative Reasoning

- Standard Deontic Logic
- Carmo and Jones logic [Carmo&Jones-2002/2012]
- Input/Output logics [Makinson&v.d.Torre-2000]
- Announcement logics

Further Reading and References:
– Universal Reasoning, Rational Argumentation and Human-Machine Interaction, `http://arxiv.org/abs/1703.09620`, 2017.
– Theorem Provers for Every Normal Modal Logic, LPAR, 2017