

## Sweet SIXTEEN

Automation via Embedding into Classical Higher-Order Logic<sup>1</sup>

Alexander Steen Christoph Benzmüller Freie Universität Berlin

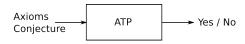
Non-Classical Logic. Theory and Applications. Seventh Edition. Torun, Poland, 2015

 $<sup>^1</sup>$ This work has been supported by the German National Research Foundation (DFG) under grant BE 2501/11-1 (Leo-III).



## Why use Automation?

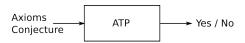
- Use automated deduction tools for assistance in verifying/refuting arguments
  - Especially interested in using higher-order ATP





## Why use Automation?

- Use automated deduction tools for assistance in verifying/refuting arguments
  - Especially interested in using higher-order ATP
- ► That should also be possible in non-classical logics
  - Feasible automation often hard





## Why use Automation?

- Use automated deduction tools for assistance in verifying/refuting arguments
  - Especially interested in using higher-order ATP
- ▶ That should also be possible in non-classical logics
  - Feasible automation often hard
- Applications in linguistics, computer science and philosophy
  - Formalization of arguments in domain-specific logic



#### Motivation



# Can anything (non-trivial) be done with it?

Yes, at least to some extent

 Some reasoning tools reached maturity and offer fair usability,
 e.g. Isabelle [NipkowPaulsenWenzel,202]



▶ Proof of Kepler's conjecture (Flyspeck project) [Hales et al., 2014]



► Formal verification/inspection of *Gödel's* ontological proof and various versions of it [BenzmüllerWoltzenlogel Paleo, ECAI 2014]



## Automation of Non-Classical Logics



#### We want

- Automation of non-classical logics
- Reasoning with uncertainty, vagueness

#### **But often**

- Proof calculi suited for automation scarce
- and so are tools

## Automation of Non-Classical Logics



#### We want

- Automation of non-classical logics
- Reasoning with uncertainty, vagueness

#### **But often**

- Proof calculi suited for automation scarce
- and so are tools

## Instead, we

- exploit the expressiveness of classical higher-order logic
- and encode the target logic within HOL



Here: Focus on SIXTEEN [ShramkoWansing,2011]

Truth-values given by

$$\mathbf{16} = 2^{2^{\{T,F\}}}$$
=  $2^{\{N,T,F,B\}}$   
=  $\{N, N, T, F, B, NT, NF, ..., NTFB\}$ 

- ► Generalization of well-known Dunn/Belnap 4-valued system
- ▶ yielding *Trilattice* (**16**,  $\sqcap_i$ ,  $\sqcup_i$ ,  $\sqcap_t$ ,  $\sqcup_t$ ,  $\sqcap_f$ ,  $\sqcup_f$ )



#### **Orders**

by information:

$$x \leq_i y : \Leftrightarrow x \subseteq y$$

by truthhood:

$$x \le_t y : \iff x^t \subseteq y^t \land y^{-t} \subseteq x^{-t}$$

by falsehood:

$$x \leq_f y : \Leftrightarrow x^f \subseteq y^f \land y^{-f} \subseteq x^{-f}$$

## **Operations**

for each (independent) axis

Meet 
$$\sqcup_{\star}$$
 Join  $\sqcap$ , Inversion  $-_{\star}$  for  $\star \in \{i, t, f\}$ 

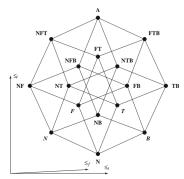


Image: Shramko, Wansing, *Truth and Falsehood* 



### Languages

 $\mathcal{L}_t$ :  $A ::= x \mid \sim_t A \mid A \land_t A \mid A \lor_t A$ 

 $\mathcal{L}_f: A ::= X \mid \sim_f A \mid A \land_f A \mid A \lor_f A$ 

 $\mathcal{L}_{tf}: A ::= X \mid \sim_t A \mid \sim_f A \mid A \land_t A \mid A \lor_t A \mid A \land_f A \mid A \lor_f A$ 



## Languages

 $\mathcal{L}_t$ :  $A := x \mid \sim_t A \mid A \land_t A \mid A \lor_t A$ 

 $\mathcal{L}_f$ :  $A := X \mid \sim_f A \mid A \land_f A \mid A \lor_f A$ 

 $\mathcal{L}_{tf}: A ::= X \mid \sim_t A \mid \sim_f A \mid A \land_t A \mid A \lor_t A \mid A \land_f A \mid A \lor_f A$ 

### Value of formulae

$$v(A \wedge_{\star} B) = v(A) \sqcap_{\star} v(B)$$
$$v(A \vee_{\star} B) = v(A) \sqcup_{\star} v(B)$$
$$v(\sim_{\star} A) = -_{\star} v(A)$$

## **Validity**

 $A \models_{\star} B : \Leftrightarrow v(A) \leq_{\star} v(B)$  for all valuations v

for  $\star \in \{t, f\}$ 



# Sadly, no deduction system is available!

▶ in fact, for many non-classical logics

#### To that end:

We embed  $\mathcal{L}_{\star}$  in classical higher-order logic (HOL).



- ► Simple types *T* generated by base types and →
- Typically, base types are o and i



Due to Alonzo Church's "Simple type theory" [Church, J.Symb.L., 1940]

- ► Simple types T generated by base types and →
- Typically, base types are o and i

Type of truth-values



Due to Alonzo Church's "Simple type theory" [Church, J.Symb.L., 1940]

- ► Simple types *T* generated by base types and →
- Typically, base types are o and i

Type of indivinduals



- ► Simple types *T* generated by base types and →
- ▶ Typically, base types are *o* and *i*
- ► Terms defined by  $(\alpha \in T)$

$$s, t := p_{\alpha} | X_{\alpha}$$



- ► Simple types *T* generated by base types and →
- ► Typically, base types are *o* and *i*
- ▶ Terms defined by  $(\alpha \in T)$

$$s, t := p_{\alpha} | X_{\alpha}$$
  
 $| (\lambda X_{\alpha}.s_{\beta})_{\alpha \to \beta} | (s_{\alpha \to \beta} t_{\alpha})_{\beta}$ 



- ► Simple types T generated by base types and →
- ► Typically, base types are *o* and *i*
- ▶ Terms defined by  $(\alpha \in T)$

```
s, t ::= p_{\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} \lor_{o \to o \to o} t_{o})_{o} \mid (\forall^{\alpha}X_{\alpha}.s_{o})_{o}
```



# Due to Alonzo Church's "Simple type theory"

[Church, J.Symb.L., 1940]

- Simple types T generated by base types and →
- Typically, base types are o and i
- ▶ Terms defined by  $(\alpha \in T)$

$$s, t ::= p_{\alpha} \mid X_{\alpha}$$

$$\mid (\lambda X_{\alpha}.s_{\beta})_{\alpha \to \beta} \mid (s_{\alpha \to \beta} t_{\alpha})_{\beta}$$

$$\mid (\neg_{o \to o}s_{o})_{o} \mid (s_{o} \lor_{o \to o \to o} t_{o})_{o} \mid \sqcap^{\alpha}(\alpha \to o) \to o \ (\lambda X_{\alpha}.s_{o})_{\alpha \to o}$$



# Due to Alonzo Church's "Simple type theory" [Church, J.Symb.L., 1940]

- ► Simple types T generated by base types and →
- Typically, base types are o and i
- ▶ Terms defined by  $(\alpha \in T)$

$$s, t ::= p_{\alpha} \mid X_{\alpha}$$

$$\mid (\lambda X_{\alpha}.s_{\beta})_{\alpha \to \beta} \mid (s_{\alpha \to \beta} t_{\alpha})_{\beta}$$

$$\mid (\neg_{o \to o}s_{o})_{o} \mid (s_{o} \lor_{o \to o \to o}t_{o})_{o} \mid (\forall^{\alpha}X_{\alpha}.s_{o})_{o}$$

Formulae of HOL are those terms with type o



- ▶  $D = \{D_{\alpha}\}_{{\alpha} \in T}$  is a frame, that is, a family of non-empty sets  $D_{\alpha}$  s.t.
  - ▶ D<sub>0</sub> = {T, F} is the set of truth and falsehood
  - ▶  $D_{\alpha \to \beta}$  is a set of functions mapping  $D_{\alpha}$  into  $D_{\beta}$
- ▶  $I = \{I_{\alpha}\}_{{\alpha} \in T}$  is a family of typed interpretation functions mapping each  $p_{\alpha}$  to an element of  $D_{\alpha}$



- ▶  $D = \{D_{\alpha}\}_{{\alpha} \in T}$  is a frame, that is, a family of non-empty sets  $D_{\alpha}$  s.t.
  - $D_0 = \{T, F\}$  is the set of truth and falsehood
  - ►  $D_{\alpha \to \beta}$  is a set of functions mapping  $D_{\alpha}$  into  $D_{\beta}$
- ▶  $I = \{I_{\alpha}\}_{{\alpha} \in T}$  is a family of typed interpretation functions mapping each  $p_{\alpha}$  to an element of  $D_{\alpha}$
- ▶ Value of a statement s denoted by  $||s||^{M,g}$  (wrt. model M and var. assignment g)



- ▶  $D = \{D_{\alpha}\}_{{\alpha} \in T}$  is a frame, that is, a family of non-empty sets  $D_{\alpha}$  s.t.
  - $D_0 = \{T, F\}$  is the set of truth and falsehood
  - ►  $D_{\alpha \to \beta}$  is a set of functions mapping  $D_{\alpha}$  into  $D_{\beta}$
- ▶  $I = \{I_{\alpha}\}_{{\alpha} \in T}$  is a family of typed interpretation functions mapping each  $p_{\alpha}$  to an element of  $D_{\alpha}$
- ▶ Value of a statement s denoted by  $||s||^{M,g}$   $\binom{\text{wrt. model } M \text{ and }}{\text{var. assignment } g}$
- ▶ *M* is a *standard model* iff  $D_{\alpha \to \beta} = D_{\beta}^{D_{\alpha}}$
- ▶ ... a Henkin model iff  $D_{\alpha \to \beta} \subseteq D_{\beta}^{D_{\alpha}}$  and  $\|.\|^{M,g}$  is a total function

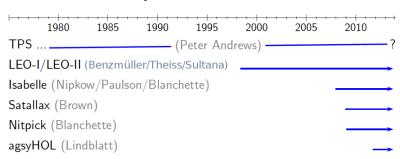


- ▶  $D = \{D_{\alpha}\}_{{\alpha} \in T}$  is a frame, that is, a family of non-empty sets  $D_{\alpha}$  s.t.
  - $D_0 = \{T, F\}$  is the set of truth and falsehood
  - ▶  $D_{\alpha \to \beta}$  is a set of functions mapping  $D_{\alpha}$  into  $D_{\beta}$
- ▶  $I = \{I_{\alpha}\}_{{\alpha} \in T}$  is a family of typed interpretation functions mapping each  $p_{\alpha}$  to an element of  $D_{\alpha}$
- ▶ Value of a statement s denoted by  $||s||^{M,g}$   $\binom{\text{wrt. model } M \text{ and }}{\text{var. assignment } g}$
- ▶ *M* is a *standard model* iff  $D_{\alpha \to \beta} = D_{\beta}^{D_{\alpha}}$
- ▶ ... a Henkin model iff  $D_{\alpha \to \beta} \subseteq D_{\beta}^{D_{\alpha}}$  and  $\|.\|^{M,g}$  is a total function
- ▶ Validity  $\models^{HOL}$  is then defined as  $M, g \models^{HOL} s_o : \Leftrightarrow ||s_o||^{M,g} = T$

#### Automation of HOL



Quite a number of systems available



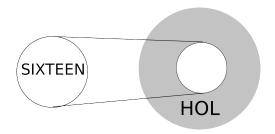
- ► TPTP THF Syntax [SutcliffeBenzmüller, J. Form. Reas., 2009]
- ► Can be called remotely via SystemOnTPTP at Miami [Sutcliffe, J. Autom. Reas., 2009]



## **Encoding of SIXTEEN in HOL**

We now encode  $(\star \in \{t, f\})$ 

- ▶ the truth-degrees of SIXTEEN
- the orderings ≤<sub>\*</sub>
- the operations □<sub>\*</sub>, □<sub>\*</sub>, −<sub>\*</sub>





Sets M are represented by their characteristic function

$$f_M(x) = \begin{cases} T & \text{, iff } x \in M \\ F & \text{otherwise} \end{cases}$$



Sets *M* are represented by their characteristic function

$$f_M(x) = \begin{cases} T & \text{, iff } x \in M \\ F & \text{otherwise} \end{cases}$$

## Set representation

Object	$\longrightarrow$	Representation in HOL
Set $M$ of objects of type $\alpha$	<b>→</b>	$f_M:\alpha\to o$
e.g. $\{x \in T \mid P(x)\}$	<b>→</b>	$\lambda x_T$ . $P_{T \to o} x$



Sets M are represented by their characteristic function

$$f_M(x) = \begin{cases} T & \text{, iff } x \in M \\ F & \text{otherwise} \end{cases}$$

## Set representation

Object
$$\longrightarrow$$
 Representation in HOLSet M of objects of type  $\alpha$  $\longrightarrow$   $f_M: \alpha \to o$ e.g.  $\{x \in T \mid P(x)\}$  $\longrightarrow$   $\lambda x_T. P_{T \to o} x$ 

## Truth-degrees as HOL sets

$$(o \rightarrow o) \rightarrow o$$



Sets M are represented by their characteristic function

$$f_M(x) = \begin{cases} T & \text{, iff } x \in M \\ F & \text{otherwise} \end{cases}$$

### Set representation

Object	$\longrightarrow$	Representation in HOL
Set $M$ of objects of type $\alpha$	<b>→</b>	$f_M: \alpha \to o$
eq $\{x \in T \mid P(x)\}$	<b>→</b>	$\lambda x_{\tau} P_{\tau} \circ x$

## Truth-degrees as HOL sets

$$(o \rightarrow o) \rightarrow 0$$



Sets M are represented by their characteristic function

$$f_M(x) = \begin{cases} T & \text{, iff } x \in M \\ F & \text{otherwise} \end{cases}$$

### Set representation

Object	$\longrightarrow$	Representation in HOL
Set $M$ of objects of type $\alpha$	<b>→</b>	$f_M:\alpha\to o$
eq $\{x \in T \mid P(x)\}$	<b>→</b>	$\lambda x_{\tau} P_{\tau} \circ x$

## Truth-degrees as HOL sets

$$(o \rightarrow o) \rightarrow o$$



Sets *M* are represented by their characteristic function

$$f_M(x) = \begin{cases} T & \text{, iff } x \in M \\ F & \text{otherwise} \end{cases}$$

## Set representation

Object	$\longrightarrow$	Representation in HOL
Set $M$ of objects of type $\alpha$	<b>→</b>	$f_M: \alpha \to o$
e.g. $\{x \in T \mid P(x)\}$	<b>→</b>	$\lambda x_T. P_{T \to o} x$

## Truth-degrees as HOL sets

$$(o \rightarrow o) \rightarrow o$$



## **Example**

Encoding of  $\mathbf{F}$  (the singleton set  $\{\mathbf{F}\}$ ):

$$\mathbf{F} \equiv \lambda n_{00}$$
.



## **Example**

Encoding of  $\mathbf{F}$  (the singleton set  $\{\mathbf{F}\}$ ):

$$\mathbf{F} \stackrel{?}{\equiv} \lambda n_{oo}$$
. n F



## **Example**

Encoding of  $\mathbf{F}$  (the singleton set  $\{\mathbf{F}\}$ ):

$$\mathbf{F} \equiv \lambda n_{oo}$$
.  $n F \wedge \neg n T$ 

# Truth-degree mapping



N	=	$\lambda n_{oo}$ . F
N	=	$\lambda n_{oo}$ . $\neg n F \land \neg n T$
T	=	$\lambda n_{oo}$ . $\neg n F \wedge n T$
F	=	$\lambda n_{oo}$ . $n F \wedge \neg n T$
В	=	$\lambda n_{oo}$ . $n F \wedge n T$
NF	=	$\lambda n_{oo}$ . $\neg n T$
NT	=	$\lambda n_{oo}$ . $\neg n F$
NB	=	$\lambda n_{oo}. (\neg n F \wedge \neg n T) \vee (n F \wedge n T)$
FT	=	$\lambda n_{oo}. (n F \wedge \neg n T) \vee (\neg n F \wedge n T)$
FB	=	λn <sub>oo</sub> . n F
ТВ	=	λn <sub>oo</sub> . n T
NFT	=	$\lambda n_{oo}$ . $\neg n F \vee \neg n T$
NFB	=	$\lambda n_{oo}$ . $n F \vee \neg n T$
NTB	=	$\lambda n_{oo}$ . $\neg n F \lor n T$
FTB	=	$\lambda n_{oo}$ . $n F \vee n T$
Α	:=	$NFTB = \lambda n_{oo}. T$

# Encoding of all 16 truth-degrees



$$\leq_t := \lambda v_{o(oo)}.\lambda w_{o(oo)}. \forall n_{oo}. \; ((v^t\; n)\supset (w^t\; n)) \wedge ((w^{-t}\; n)\supset (v^{-t}\; n))$$



$$\leq_t := \lambda V_{o(oo)}.\lambda W_{o(oo)}.\forall n_{oo}. \underbrace{((v^t n) \supset (w^t n))}_{v^t \subseteq w^t} \wedge ((w^{-t} n) \supset (v^{-t} n))$$



$$\leq_t := \lambda v_{o(oo)}.\lambda w_{o(oo)}.\forall n_{oo}.\left((v^t\ n)\supset (w^t\ n)\right) \land \underbrace{\left((w^{-t}\ n)\supset (v^{-t}\ n)\right)}_{w^{-t}\subseteq v^{-t}}$$



$$\leq_{t} := \lambda V_{o(oo)}.\lambda W_{o(oo)}.\forall n_{oo}. ((v^{t} n) \supset (w^{t} n)) \land ((w^{-t} n) \supset (v^{-t} n))$$

$$\sqcup_{t} := \lambda V_{o(oo)}.\lambda W_{o(oo)}.v^{t} \cup w^{t} \cup (w^{-t} \cap v^{-t})$$

$$\sqcap_{t} := \lambda V_{o(oo)}.\lambda W_{o(oo)}.v^{-t} \cup w^{-t} \cup (w^{t} \cap v^{t})$$



$$\leq_t := \lambda V_{o(oo)}.\lambda W_{o(oo)}.\forall n_{oo}. ((v^t n) \supset (w^t n)) \land ((w^{-t} n) \supset (v^{-t} n))$$
 
$$\sqcup_t := \lambda V_{o(oo)}.\lambda W_{o(oo)}.v^t \cup w^t \cup (w^{-t} \cap v^{-t})$$
 
$$\sqcap_t := \lambda V_{o(oo)}.\lambda W_{o(oo)}.v^{-t} \cup w^{-t} \cup (w^t \cap v^t)$$

## with auxillary definitions:

$$(v)_{o(oo)}^t := \lambda n_{oo}. \ (v \ n) \land (n \ \mathsf{T}) \qquad (v)_{o(oo)}^{-t} := \lambda n_{oo}. \ (v \ n) \land \neg (n \ \mathsf{T})$$



# **Embedding a proof task**

Input:  $x \wedge_t y \models_t x$ 

# Proof task example



## **Embedding a proof task**

Input:

$$X \wedge_t y \models_t X$$

Translation to a HOL proof task:

$$[x \wedge_t y] [\leq_t] [x]$$

# Proof task example



### **Embedding a proof task**

Input:

$$X \wedge_t y \models_t X$$

Translation to a HOL proof task:

$$[X \wedge_t y] [\leq_t] [X]$$

Expansion of definitions,  $\beta$ -normalization:

$$\underbrace{x^{-t} \cup y^{-t} \cup (x^t \cap y^t)}_{=:A} \lceil \leq_t \rceil x$$



 $X \wedge_t y \models_t X$ 

### **Embedding a proof task**

Input:

Translation to a HOL proof task: 
$$[x \land_t y] [\leq_t] [x]$$

Expansion of definitions,  $\beta$ -normalization:

$$\forall n_{oo}. ((\lceil A^t \rceil n) \supset (\lceil x^t \rceil n)) \land ((\lceil x^{-t} \rceil n) \supset (\lceil A^{-t} \rceil n))$$



 $X \wedge_t V \models_t X$ 

#### **Embedding a proof task**

Input:

Translation to a HOL proof task: 
$$[x \land_t y] [\leq_t] [x]$$

Expansion of definitions,  $\beta$ -normalization:

$$\underbrace{x^{-t} \cup y^{-t} \cup \left(x^t \cap y^t\right)}_{=:A} \left[ \leq_t \right] x$$

$$\forall n_{oo.} \left( \left( \left[ A^t \right] n \right) \supset \left( \left[ x^t \right] n \right) \right) \wedge \left( \left( \left[ x^{-t} \right] n \right) \supset \left( \left[ A^{-t} \right] n \right) \right)$$

$$\forall n_{oo}. ((\lambda n_{oo}. (\lceil A \rceil n) \land (n \top) n) \supset (\lambda n_{oo}. (x n) \land (n \top) n)) \\ \land ((\lambda n_{oo}. (x n) \land \neg (n \top) n) \supset (\lambda n_{oo}. (\lceil A \rceil n) \land \neg (n \top) n)) \\ \downarrow$$



 $X \wedge_t V \models_t X$ 

#### **Embedding a proof task**

Input:

Translation to a HOL proof task: 
$$[x \land_t y] [\leq_t] [x]$$

Expansion of definitions,  $\beta$ -normalization:

$$\underbrace{x^{-t} \cup y^{-t} \cup (x^t \cap y^t)}_{=:A} [\leq_t] x$$

$$\forall n_{oo}. ((\lceil A^t \rceil n) \supset (\lceil x^t \rceil n)) \land ((\lceil x^{-t} \rceil n) \supset (\lceil A^{-t} \rceil n))$$

$$\forall n_{oo}. ((\lambda n_{oo}. (\lceil A \rceil n) \land (n \top) n) \supset (\lambda n_{oo}. (x n) \land (n \top) n))$$

$$\land ((\lambda n_{oo}. (x n) \land \neg (n \top) n) \supset (\lambda n_{oo}. (\lceil A \rceil n) \land \neg (n \top) n))$$

# Can now be passed to standard HOL ATP

# **Experiments**



- Embedding successfully employed
- ► Encoded in TPTP THF and Isabelle/HOL
- Some meta-logical properties verified

Example: Verfication of

**Prop. 3.2** (from Shramko and Wansing, Truth and Falsehood):

$$\mathbf{B} \in S \land \mathbf{B} \in T \Leftrightarrow \mathbf{B} \in S \sqcup T$$
 9 ms  
 $\mathbf{F} \in S \lor \mathbf{F} \in T \Leftrightarrow \mathbf{F} \in S \sqcup T$  8 ms  
 $\vdots$   $\vdots$   $\Sigma 131 \text{ ms}$ 

# **Experiments**



- Embedding successfully employed
- ▶ Encoded in TPTP THF and Isabelle/HOL
- Some meta-logical properties verified

Example: Verfication of

Prop. 3.4 (from Shramko and Wansing, Truth and Falsehood):

$$\mathbf{B} \in -_t S \Leftrightarrow \mathbf{F} \in S$$
 8 ms  
 $\mathbf{F} \in -_t S \Leftrightarrow \mathbf{B} \in S$  9 ms  
 $\vdots$   $\vdots$   $\Sigma$  69 ms

# **Experiments**



- Embedding successfully employed
- ▶ Encoded in TPTP THF and Isabelle/HOL
- Some meta-logical properties verified

Example: Verfication of

**Def. 3.6** (from Shramko and Wansing, Truth and Falsehood):

$$S \leq_t T \Rightarrow -_t T \leq_t -_t S$$
 421 ms  
 $S \leq_f T \Rightarrow -_t S \leq_f -_t T$  422 ms  
 $\vdots$   $\vdots$   $\Sigma$  1734 ms

#### Conclusion and Outlook



#### Conclusion

- embedding of SIXTEEN into HOL
- currently, no automated deduction systems available
- embedding allows automation using common HOL ATPs
- and using interactive proof assistants (e.g. Coq)
- ... including reasoning about SIXTEEN



#### Conclusion

- embedding of SIXTEEN into HOL
- currently, no automated deduction systems available
- embedding allows automation using common HOL ATPs
- and using interactive proof assistants (e.g. Coq)
- ... including reasoning about SIXTEEN

# **Outlook: Towards HOL based universal reasoning**

- augment SIXTEEN embedding with quantification (of arbitrary order?)
- many quantified non-classical logics are fragments of HOL
- logic combinations: e.g. SIXTEEN with modalities
- out-of-the-box automation via HOL ATPs for free