Automating Quantified Conditional Logics in HOL

Christoph Benzmüller

Freie Universität Berlin

Supported by the German Research Foundation (grant BE2501/9-1)

Abstract

A notion of **quantified conditional logics (QCLs)** is provided that includes quantification over **individual and propositional variables**. The former is supported with respect to **constant and variable domain semantics**. In addition, a **sound and complete embedding of this framework in classical higher-order logic (HOL)** is presented. Using prominent examples from the literature it is demonstrated how this embedding enables **effective automation of reasoning** within (object-level) and about (meta-level) quantified conditional logics with **off-the-shelf higher-order theorem provers and model finders**.

Overall Motivation and Contribution

QCLs are very expressive non-classical logics; they have many applications; no provers have been available so far. However,

- **QCLs are fragments of HOL (with Henkin semantics)**
- **and they can easily be automated as such,**
- **they inherit important meta- resp. proof-theoretical properties (cut-elimination, compactness, etc.), and**
- **they can easily be combined with other logics in HOL.**

This reserach is part of a larger project which takes HOL as starting point for studying classical and non-classical logics and their combinations.

Reading: [Benzmüller'13]

Quantified Conditional Logics (QCLs)

$$\varphi, \psi ::= \; P \mid k(X^1, \ldots, X^n) \mid \neg\varphi \mid \varphi \vee \psi \mid \varphi \Rightarrow \psi \mid \forall^{co} X \varphi \mid \forall^{va} X \varphi \mid \forall P \varphi$$

Interpretation: $M = \langle S, f, D, D', Q, I \rangle$ where S is a set of 'worlds', $f : S \times 2^S \mapsto 2^S$ is the selection function, $D \neq \emptyset$ is a set of *individuals* (constant domain), D' is a function that assigns a subset $D'(w) \neq \emptyset$ of D to each world w (varying domains), $Q \neq \emptyset$ is a collection of subsets of W (prop. domain), and I is an interpretation function s.t. for each predicate symbol k, $I(k, w) \subseteq D^n$.

Satisfiability of φ (denoted as $M, g, s \models \varphi$) for an interpretation M, a world $s \in S$, and a variable assignment $g = (g^i, g^p)$:

$$\begin{array}{l}
M, g, s \models k(X^1, \ldots, X^n) \text{ iff } \langle g^i(X^1), \ldots, g^i(X^n) \rangle \in I(k, w) \\
M, g, s \models P \text{ iff } s \in g^p(P) \\
M, g, s \models \neg\varphi \text{ iff } M, g, s \not\models \varphi \text{ (that is, not } M, g, s \models \varphi) \\
M, g, s \models \varphi \vee \psi \text{ iff } M, g, s \models \varphi \text{ or } M, g, s \models \psi \\
M, g, s \models \forall^{co} X \varphi \text{ iff } M, ([d/X]g^i, g^p), s \models \varphi \text{ for all } d \in D \\
M, g, s \models \forall^{va} X \varphi \text{ iff } M, ([d/X]g^i, g^p), s \models \varphi \text{ for all } d \in D'(s) \\
M, g, s \models \forall P \varphi \text{ iff } M, (g^i, [p/P]g^p), s \models \varphi \text{ for all } p \in Q \\
M, g, s \models \varphi \Rightarrow \psi \text{ iff } M, g, t \models \psi \text{ for all } t \in S \text{ s.t. } t \in f(s, [\varphi]) \\
\qquad \text{where } [\varphi] = \{u \mid M, g, u \models \varphi\}
\end{array}$$

$M \models^{QCL} \varphi$ iff $M, g, s \models \varphi$ for all s, g. $\models \varphi$ iff $M \models^{QCL} \varphi$ for all M.

Reading: [Stalnaker'68],[Delgrande'98]

Classical Higher-order Logic (HOL)

Types $\quad \alpha, \beta ::= \iota\,(\textit{worlds}) \mid \mu\,(\textit{indiv.}) \mid o\,(\textit{Booleans}) \mid \alpha \to \beta$

$$\begin{array}{l}
s, t ::= c_\alpha \mid X_\alpha \mid (\lambda X_\alpha s_\beta)_{\alpha\to\beta} \mid (s_{\alpha\to\beta}\, t_\alpha)_\beta \mid \\
\qquad (\neg_{o\to o}\, s_o)_o \mid (s_o \vee_{o\to o\to o} t_o)_o \mid (\Pi_{(\alpha\to o)\to o}\, s_{\alpha\to o})_o
\end{array}$$

Note: Binder notation $\forall X_\alpha t_o$ as syntactic sugar for $\Pi_{(\alpha\to o)\to o} \lambda X_\alpha t_o$

Frame: collection $\{D_\alpha\}_{\alpha\in T}$ s.t. $D_o = \{T, F\}$, $D_\iota \neq \emptyset$ and $D_u \neq \emptyset$ arbitrary, and $D_{\alpha\to\beta}$ are collections of total functions from D_α to D_β.

Interpretation: Tuple $\langle \{D_\alpha\}_{\alpha\in T}, I \rangle$ where $\{D_\alpha\}_{\alpha\in T}$ is a frame and where function I maps each typed constant symbol c_α to an appropriate element of D_α, called the *denotation* of c_α. The denotations of $\neg$, $\vee$ and $\Pi_{(\alpha\to o)\to o}$ are always chosen as usual.

An interpretation is a **Henkin model** iff there is a valuation function V s.t. $V(\phi, s_\alpha) \in D_\alpha$ for each variable assignment ϕ and term s_α, and the following conditions are satisfied: $V(\phi, X_\alpha) = \phi(X_\alpha)$, $V(\phi, c_\alpha) = I(c_\alpha)$, $V(\phi, l_{\alpha\to\beta}\, r_\alpha) = (V(\phi, l_{\alpha\to\beta}) V(\phi, r_\alpha))$, and $V(\phi, \lambda X_\alpha s_\beta)$ represents the function from D_α into D_β whose value for each argument $z \in D_\alpha$ is $V(\phi[z/X_\alpha], s_\beta)$. If an interpretation is an Henkin model the function V is uniquely determined.

$H \models^{HOL} s$ iff $V(\phi, s) = T$ for all ϕ. $\models s$ iff $H \models^{HOL} s$ for all H.

Reading: [Church'40],[Andrews'72a/b],[BenzmüllerEtAl'04]

Embedding QCLs in HOL — In other words: QCLs are simple Fragments of HOL!

The mapping $\lfloor\cdot\rfloor$ identifies QCL formulas φ with HOL terms $\lfloor\varphi\rfloor$ of type $\tau := \iota \to o$. The mapping is recursively defined:

$$\begin{array}{lcl}
\lfloor P \rfloor & = & P_\tau \\
\lfloor k(X^1, \ldots, X^n) \rfloor & = & k_{u^n\to\tau}\, X_u^1 \ldots X_u^n \\
\lfloor \neg\varphi \rfloor & = & \neg_{\tau\to\tau} \lfloor\varphi\rfloor \\
\lfloor \varphi \vee \psi \rfloor & = & \vee_{\tau\to\tau\to\tau} \lfloor\varphi\rfloor \lfloor\psi\rfloor \\
\lfloor \varphi \Rightarrow \psi \rfloor & = & \Rightarrow_{\tau\to\tau\to\tau} \lfloor\varphi\rfloor \lfloor\psi\rfloor \\
\lfloor \forall^{co} X \varphi \rfloor & = & \Pi^{co}_{(u\to\tau)\to\tau}\, \lambda X_u \lfloor\varphi\rfloor \\
\lfloor \forall^{va} X \varphi \rfloor & = & \Pi^{va}_{(u\to\tau)\to\tau}\, \lambda X_u \lfloor\varphi\rfloor \\
\lfloor \forall P \varphi \rfloor & = & \Pi_{(\tau\to\tau)\to\tau}\, \lambda P_\tau \lfloor\varphi\rfloor
\end{array}$$

P_τ and $X_u^1, \ldots, X_u^n$ are variables and $k_{u^n\to\tau}$ is a constant symbol.

$\neg_{\tau\to\tau}$, $\vee_{\tau\to\tau\to\tau}$, $\Rightarrow_{\tau\to\tau\to\tau}$, $\Pi^{co,va}_{(u\to\tau)\to\tau}$ and $\Pi_{(\tau\to\tau)\to\tau}$ realize the QCL connectives in HOL. They abbreviate the following HOL terms:

$$\begin{array}{lcl}
\neg_{\tau\to\tau} & = & \lambda A_\tau \lambda X_\iota \neg(A\, X) \\
\vee_{\tau\to\tau\to\tau} & = & \lambda A_\tau \lambda B_\tau \lambda X_\iota (A\, X \vee B\, X) \\
\Rightarrow_{\tau\to\tau\to\tau} & = & \lambda A_\tau \lambda B_\tau \lambda X_\iota \forall V_\iota (f\, X\, A\, V \to B\, V) \\
\Pi^{co}_{(u\to\tau)\to\tau} & = & \lambda Q_{u\to\tau} \lambda V_\iota \forall X_u (Q\, X\, V) \\
\Pi^{va}_{(u\to\tau)\to\tau} & = & \lambda Q_{u\to\tau} \lambda V_\iota \forall X_u (eiw\, V\, X \to Q\, X\, V) \\
\Pi_{(\tau\to\tau)\to\tau} & = & \lambda R_{\tau\to\tau} \lambda V_\iota \forall P_\tau (R\, P\, V)
\end{array}$$

The interpretations of f and eiw are chosen appropriately. For the varying domains non-emptiness is postulated: $\forall W_\iota \exists X_u (eiw\, W\, X)$

Meta-level notion of validity defined as $\mathrm{vld}_{\tau\to o} = \lambda A_\tau \forall S_\iota (A\, S)$.

Theorem: Soundness and Completeness

$\models^{QCL} \varphi \;$ iff $\; \{NE\} \models^{HOL} \mathrm{vld}_{\tau\to o} \lfloor\varphi\rfloor$ (wrt Henkin semantics)

(Proof: By relating Kripke structures to Henkin models.)

Corollary: Cut-elimination for QCL

There are cut-free calculi for QCL.

(Proof: Take any cut-free calculus for HOL, e.g. the cut-free sequent calculus from [BenzmüllerEtAl'09]. Note, however, the potential impact of cut-simulation.)

Reading: Earlier work is reported in [Benzm.Genovese'11]

The Encoding in THF0-Syntax

```
%---- file: Axioms.ax ------------------------------------
%--- type mu for individuals
thf(mu,type,(mu:$tType)).
%--- reserved constant for selection function f
thf(f,type,(f:$i>($i>$o)>$i>$o)).
%--- `exists in world' predicate for varying domains;
%--- for each v we get a non-empty subdomain eiw@v
thf(eiw,type,(eiw:$i>mu>$o)).
thf(nonempty,axiom,(![V:$i]:?[X:mu]:(eiw@V@X))).
%--- negation, disjunction, material implication
thf(not,type,(not:($i>$o)>$i>$o)).
thf(or,type,(or:($i>$o)>($i>$o)>$i>$o)).
thf(impl,type,(impl:($i>$o)>($i>$o)>$i>$o)).
thf(not_def,definition,(not = (^[A:$i>$o,X:$i]:~(A@X)))).
thf(or_def,definition,(or
 = (^[A:$i>$o,B:$i>$o,X:$i]:((A@X)|(B@X))))).
thf(impl_def,definition,(impl
 = (^[A:$i>$o,B:$i>$o,X:$i]:((A@X)=>(B@X))))).
%--- conditionality
thf(cond,type,(cond:($i>$o)>($i>$o)>$i>$o)).
thf(cond_def,definition,(cond
 = (^[A:$i>$o,B:$i>$o,X:$i]:![W:$i]:((f@X@A@W)=>(B@W))))).
%--- quantification (constant dom., varying dom., prop.)
thf(all_co,type,(all_co: (mu>$i>$o)>$i>$o)).
thf(all_va,type,(all_va:(mu>$i>$o)>$i>$o)).
thf(all,type,(all:(($i>$o)>$i>$o)>$i>$o)).
thf(all_co_def,definition,(all_co
 = (^[A:mu>$i>$o,W:$i]:![X:mu]:(A@X@W)))).
thf(all_va_def,definition,(all_va
 = (^[A:mu>$i>$o,W:$i]:![X:mu]:((eiw@W@X)=>(A@X@W))))).
thf(all_def,definition,(all
 = (^[A:($i>$o)>$i>$o,W:$i]:![P:$i>$o]:(A@P@W)))).
%--- box operator based on conditionality
thf(box,type,(box:($i>$o)>$i>$o)).
thf(box_def,definition,(box
 = (^[A:$i>$o]:(cond@(not@A)@A)))).
%--- notion of validity of a conditional logic formula
thf(vld,type,(vld:($i>$o)>$o)).
thf(vld_def,definition,(vld
 = (^[A:$i>$o]:![S:$i]:(A@S)))).
%---- end file: Axioms.ax --------------------------------
```

Reading: Introduction to THF0-Syntax [SutcliffeBenzm.'10]

Automating Prominent Examples from the Literature (in QCL+ID+MP)

Example: Pegasus, the winged horse

It can be consistently stated (in QCL+ID+MP) that: *"Horses (h) contingently do not have wings (w) but Pegasus (p) is a winged horse."*

$$\forall^{va} X (h(X) \to \neg w(X)), \quad h(p), \quad w(p)$$

THF0 encoding of this example:

```
%----------------------------------------------
include('Axioms.ax').
%--- axioms ID and MP
thf(id,axiom,
 (vld@(all@^[P:$i>$o]:(cond@P@P)))).
thf(mp,axiom,
 (vld@(all@^[P:$i>$o]:(all@^[Q:$i>$o]:
  (impl@(cond@P@Q)@(impl@P@Q)))))).
%--- type declarations
thf(horse,type,(horse:mu>$i>$o)).
thf(wings,type,(wings:mu>$i>$o)).
thf(fly,type,(fly:mu>$i>$o)).
thf(pegasus,type,(pegasus:mu)).
%--- the statements
thf(ax1,axiom,
 (vld@(all_va@^[X:mu]:
  (impl@(horse@X)@(not@(wings@X)))))).
thf(ax2,axiom,(vld@(horse@pegasus))).
thf(ax3,axiom,(vld@(wings@pegasus))).
%----------------------------------------------
```

H confirms the satisfiability of these formulas (with H_N=7.7). The finite model generated by Nitpick tells us that Pegasus is not 'actual', i.e., does not exist (cf. eiw) in any world. As expected, when the example problem is formulated with $\forall^{co}$ instead of $\forall^{va}$ then H reports unsatisfiability ($H_{L,S}$=0.0, H_I=5.8).

Notation: $\phi \Rightarrow_X \psi := (\exists^{va} X \phi) \Rightarrow \forall^{va} X (\phi \to \psi)$

Example: Opus, the penguin

"Birds (b) normally fly (f), but Opus (o) is a bird that normally does not fly."

$$b(X) \Rightarrow_X f(X), \quad b(o), \quad b(o) \Rightarrow \neg f(o)$$

H reports a finite model (H_N=8.6). When $\forall^{co}$ is used: H says unsatisfiable (H_S=0.0, H_I=7.9).

"Birds normally fly and necessarily Opus the bird does not fly."

$$b(X) \Rightarrow_X f(X), \quad \Box(b(o) \wedge \neg f(o))$$

H reports a finite model (H_N=8.7). When $\forall^{co}$ is used: H says unsatisfiable (H_S=0.0, H_I=7.6).

"Birds normally fly and necessarily there is a non-flying bird."

$$b(X) \Rightarrow_X f(X), \quad \Box\exists^{va}(b(X) \wedge \neg f(X))$$

H reports unsatisfiability (H_S=0.0, H_I=8.7), also when $\forall^{co}$ is used (H_S=0.0, H_I=8.8).

"Birds normally fly, penguins normally do not fly and that all penguins are necessarily birds."

$$b(X) \Rightarrow_X f(X), \quad p(X) \Rightarrow_X \neg f(X), \quad \forall^{va}\Box(p(X) \to b(X))$$

H generates a finite model ($\forall^{va}$: H_N=8.8; $\forall^{co}$: H_N=7.9).

Moreover, H can conclude from the statements above that *"Birds are normally not penguins."* ($\forall^{va}$: H_S=0.9, H_L=10.2, H_A=9.4; $\forall^{co}$: H_S=0.8, H_L=10.1, H_A=0.3):

$$b(X) \Rightarrow_X f(X), \quad p(X) \Rightarrow_X \neg f(X), \quad \forall^{va}\Box(p(X) \to b(X)) \quad \vdash \quad b(X) \Rightarrow_X \neg p(X)$$

In line with Delgrande, H reports a countermodel for the following statement (H_N=8.7):

$$b(X) \Rightarrow_X f(X), \quad p(X) \Rightarrow_X \neg f(X), \quad \forall^{va}\Box(p(X) \to b(X)) \quad \vdash \quad b(o) \Rightarrow \neg p(o)$$

However, when $\forall^{co}$ is used, H reports a theorem (H_S=0.8, H_A=0.4).

Reading: These examples have been discussed (but not automated) in [Delgrande'98]

The HOL Metaprover *H*

The H metaprover for HOL sequentially calls the following prover and model finders:

- H_L LEO-II (Benzmüller/Sultana/Theiss): `http://www.leoprover.org`
- H_S Satallax (Brown): `http://www.ps.uni-saarland.de/~cebrown/satallax/`
- H_I Isabelle (Blanchette/Paulson/Nipkow/…): `http://isabelle.in.tum.de/`
- H_N Nitpick (Blanchette): `http://www4.in.tum.de/~blanchet/nitpick.html`
- H_A agsyHol (Lindblatt): `https://github.com/frelindb/agsyHOL`

These systems support THF0 syntax. These provers are remotely available via SystemOnTPTP: `http://www.cs.miami.edu/~tptp/cgi-bin/SystemOnTPTP`

References and Further Reading

[Andrews'72a] P.B. Andrews. General models, descriptions, and choice in type theory. JSL, 37(2):385-394, 1972.
[Andrews'72b] P.B. Andrews. General models and extensionality. JSL, 37(2):395–397, 1972.
[Benzmüller'13] C. Benzmüller. A top-down approach to combining logics, Proc. of ICAART 2013, Barcelona, Spain, 2013.
[BenzmüllerEtAl'04] C. Benzmüller, C. E. Brown, and M. Kohlhase, Higher order semantics and extensionality. JSL, 69(4):1027-1088, 2004.
[BenzmüllerEtAl'09] C. Benzmüller, C. E. Brown, and M. Kohlhase. Cut-simulation and impredicativity. LMCS, 5(1:6):1–21, 2009.
[Benzm.Genovese'11] C. Benzmüller and V. Genovese. Quantified conditional logics are fragments of HOL. NCMPL 2011. arXiv:1204.5920v1
[Church'40] A. Church. A formulation of the simple theory of types. JSL, 5:56–68, 1940.
[Delgrande'98] J.P. Delgrande. On first-order conditional logics. Artificial Intelligence, 105(1-2):105–137, 1998.
[Stalnaker'68] R.C. Stalnaker. A theory of conditionals. In Studies in Logical Theory, pp. 98–112. Blackwell, 1968.
[SutcliffeBenzm.'10] G. Sutcliffe and C. Benzm. Automated reasoning in HOL using the TPTP THF infrastructure. *JFR*, 3(1):1–27, 2010.