On a (quite) universal theorem proving approach and its application in metaphysics

Christoph Benzmüller*

Freie Universität Berlin, Germany
c.benzmueller@fu-berlin.de

Abstract. Classical higher-order logic is suited as a meta-logic in which a range of other logics can be elegantly embedded. Interactive and automated theorem provers for higher-order logic are therefore readily applicable. By employing the approach the automation of a variety of ambitious logics has recently been pioneered, including variants of first-order and higher-order quantified multimodal logics and conditional logics. Moreover, the approach supports the automation of meta-level reasoning, and it sheds some new light on meta-theoretical results such as cut-elimination. Most importantly, however, the approach is relevant for practice: it has recently been successfully applied in a series of experiments in metaphysics in which higher-order theorem provers have actually contributed some new knowledge.

In 2008, in a collaboration with Larry Paulson, I have started to study embeddings of first-order and higher-order quantified multimodal logics in classical higher-order logic (HOL) [15, 17]. Key motivation has been the automation of non-classical logics for which no automated theorem provers (ATPs) were available till then. Together with colleagues and students the approach has since been further developed and adapted for a range of other non-classical logics [16, 3, 10, 19, 2, 12, 6, 4, 20, 22, 9, 8, 40]. A recent highlight has been the application of the approach to a prominent and widely discussed argument in metaphysics: Kurt Gödel's ontological argument for the existence of God [14, 13]. This work, conducted jointly with Bruno Woltzenlogel Paleo (TU Vienna, Austria; now ANU Canberra, Australia), received a media repercussion on a global scale. The logic embedding approach has been central to this success.

Section 1 outlines the main advantages of the approach, and Section 2 discusses some key results from our application studies in metaphysics.

1 Advantages of the logic embedding approach

Pragmatics and convenience. 'Implementing' an interactive or automated theorem prover is made very simple, even for very challenging quantified non-classical

* This work has been supported by the German Research Foundation DFG under grants BE2501/9-1,2 and BE2501/11-1.

logics. The core idea is to introduce the connectives (and meta-level predicates such as 'validity') of the embedded logic as abbreviations of certain lambda terms in HOL, for example, by encoding Kripke style semantics. Exemplary embeddings for various challenging logics have been discussed in the papers referenced above. Amongst these logics are variants of conditional logic, multimodal logic, intuitionistic logic, hybrid logic, tense logic, paraconsistent logic, etc. For the mentioned application in metaphysics it is has been particularly important to mechanise variants of higher-order modal logics (HOML).

Flexibility. The approach is flexible and supports rapid experimentations with logic variations. For example, quantifiers for constant, varying and cumulative domains may be introduced, rigid or non-rigid terms may be considered. Moreover, in order to arrive at particular modal logics such as S4 or S5 from base logic K, respective Sahlqvist axioms may be postulated. Alternatively (and preferably), one may simply state the corresponding conditions (like symmetry, reflexivity and transitivity) of the accessibility relation directly in HOL. Analogous logic axiomatisations are possible for e.g. conditional logics. Moreover, to support multiple modalities, indexed box operators (the indices being accessibility relations) can be formalised and different combination schemes are possible. Furthermore, prominent connections between logics can be formalised and exploited. For example, Fig. 2 in [18] shows how the modal $\Box$-operator can be defined in terms of conditional implication.

Availability. The embedding approach is readily available. Option one is to reuse and adapt the TPTP THF0 [39] encodings of the various logic embeddings as provided in our papers (see e.g. Fig.1 in [22]). This turns any THF0-compliant prover, such as LEO-II [18], Satallax[27] or Nitpick [26], into a reasoner for the embedded logic. Note that a range of prominent THF0 provers can even be accessed remotely via Geoff Sutcliffe's SystemOnTPTP infrastructure [38]. Options two and three are to reuse and adapt our Isabelle [32] and Coq [24] encodings (see e.g. Sections 4.2 and 4.3 in [22]). This turns these prominent systems into proof assistants for the embedded logics, and tools like Sledgehammer [25] can be employed to call external HOL ATPs. In many experiments we have even employed these three options simultaneously.

Relation to labelled deductive systems. The embedding approach is related to labelled deductive systems [29], which employ meta-level (world-)labelling techniques for the modeling and implementation of non-classical proof systems. In the embedding approach such labels are instead encoded directly in the HOL logic; no extra-logical annotations are required.

Relation to the standard translation. The embedding of modal logics in our approach is related to the standard relational translation [33]. In fact, (for propositional modal logics) the approach can be seen as intra-logical formalisation and implementation of the standard translation in terms of a set of (equational) axioms or definitions in HOL. However, in our work we have extended the approach to various other logics, and, in particular, to support first-order and higher-order quantification including different domain conditions. Future work could investigate whether the functional translation [34] could provide a suitable alternative to the current relational core of the approach.

Soundness and completeness. The embedding approach has been shown sound and complete for a range of different logics, see e.g. [17, 4, 14]. The reference semantics for HOL has been Henkin semantics, that is, the semantics that is also supported by THF0 compliant higher-order provers [11].

Meta-reasoning. Reasoning about logics and about logic relationships is supported in the embedding approach. For example, a systematic verification of the modal logic cube in Isabelle is presented in [9] and Fig. 10 in [22] illustrates the verification of some meta-level results on description logic ALC (soundness of the usual ALC tableaux rules and correspondence between ALC and base modal logic K). Some meta-level results for conditional logics are presented in [22].

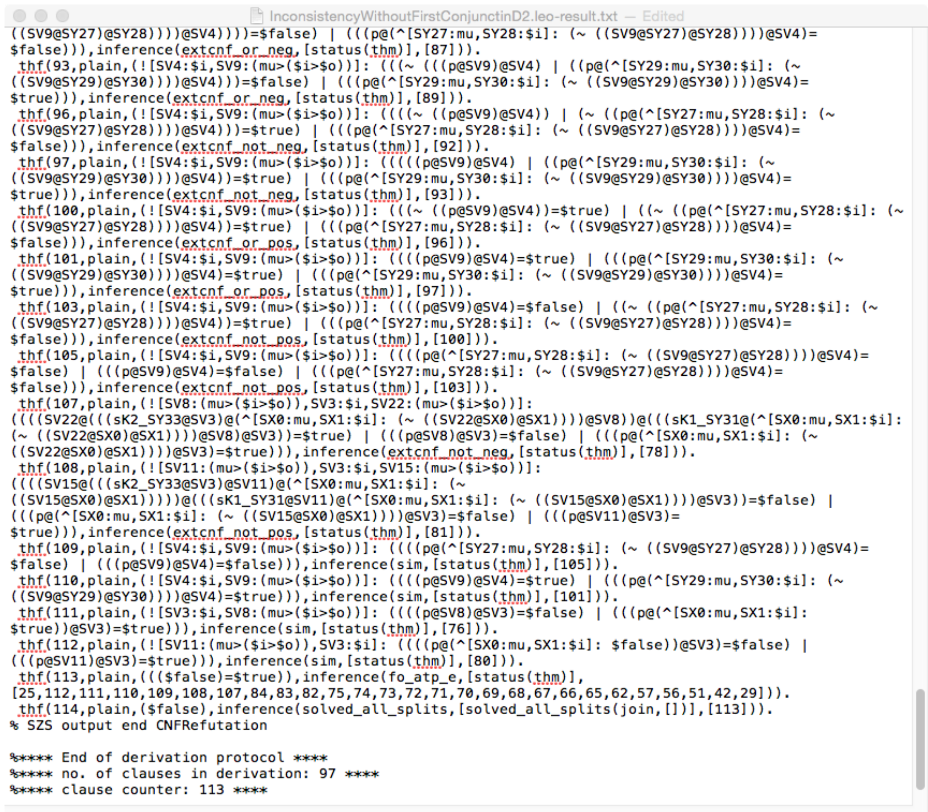
Cut-elimination. At a proof-theoretic level, the approach gives rise to a very generic (but indirect) cut-elimination result for the embedded logics [5]. This work combines the soundness and completeness results mentioned above with the fact that HOL already enjoys cut-elimination for Henkin semantics [7].

Direct calculi and user intuition. The approach supports the additional implementation of 'direct' proof calculi on top of the respective logic embeddings. For example, in [23] the implementation of a natural deduction style calculus for HOML in Coq is presented; the rules of this calculus are modeled as abstract-level tactics on top of the underlying embedding of HOML in Coq. Human intuitive proofs are thereby enabled at the interaction layer, and proofs developed at that level are directly verified by expanding the embedding in HOL. Automation attempts with HOL ATPs can be handled as before. The combination of the direct approach and the embedding approach thus provides an interesting perspective for mixed proof developments. Future work could also investigate whether proof planning [31, 28] can be employed to additionally automate the abstract-level direct proof calculi. Proof assistants in the style of *Ω*mega [36] could eventually be adapted for this, and *Ω*mega's support for 3-dimensional proof objects might turn out particularly useful in this context.

2 Results from recent applications in metaphysics

In recent work [14, 13] we have applied the embedding approach to investigate a philosophical argument that has fascinated philosophers and theologians for about 1000 years: the ontological argument for the existence of God [37].

Our initial focus was on Gödel's [30] modern version of this argument (which is in the tradition of the work of Anselm of Canterbury) and on Scott's [35] modification. Both employ a second-order modal logic (S5) for which, until now, no



```
InconsistencyWithoutFirstConjunctinD2.leo-result.txt — Edited
((SV9@SY27)@SY28))))@SV4))))=$false) | (((p@(^[SY27:mu,SY28:$i]: (~ ((SV9@SY27)@SY28))))@SV4)=
$false))),inference(extcnf_or_neg,[status(thm)],[87])).
 thf(93,plain,(![SV4:$i,SV9:(mu>($i>$o))]: (((~ (((p@SV9)@SV4) | ((p@(^[SY29:mu,SY30:$i]: (~
((SV9@SY29)@SY30))))@SV4)))=$false) | (((p@(^[SY29:mu,SY30:$i]: (~ ((SV9@SY29)@SY30))))@SV4)=
$true))),inference(extcnf_or_neg,[status(thm)],[89])).
 thf(96,plain,(![SV4:$i,SV9:(mu>($i>$o))]: ((((~ ((p@SV9)@SV4)) | (~ ((p@(^[SY27:mu,SY28:$i]: (~
((SV9@SY27)@SY28))))@SV4)))=$true) | (((p@(^[SY27:mu,SY28:$i]: (~ ((SV9@SY27)@SY28))))@SV4)=
$false))),inference(extcnf_not_neg,[status(thm)],[92])).
 thf(97,plain,(![SV4:$i,SV9:(mu>($i>$o))]: (((((p@SV9)@SV4) | ((p@(^[SY29:mu,SY30:$i]: (~
((SV9@SY29)@SY30))))@SV4))=$true) | (((p@(^[SY29:mu,SY30:$i]: (~ ((SV9@SY29)@SY30))))@SV4)=
$true))),inference(extcnf_not_neg,[status(thm)],[93])).
 thf(100,plain,(![SV4:$i,SV9:(mu>($i>$o))]: (((~ ((p@SV9)@SV4))=$true) | ((~ ((p@(^[SY27:mu,SY28:$i]: (~
((SV9@SY27)@SY28))))@SV4))=$true) | (((p@(^[SY27:mu,SY28:$i]: (~ ((SV9@SY27)@SY28))))@SV4)=
$false))),inference(extcnf_or_pos,[status(thm)],[96])).
 thf(101,plain,(![SV4:$i,SV9:(mu>($i>$o))]: ((((p@SV9)@SV4)=$true) | (((p@(^[SY29:mu,SY30:$i]: (~
((SV9@SY29)@SY30))))@SV4)=$true) | (((p@(^[SY29:mu,SY30:$i]: (~ ((SV9@SY29)@SY30))))@SV4)=
$true))),inference(extcnf_or_pos,[status(thm)],[97])).
 thf(103,plain,(![SV4:$i,SV9:(mu>($i>$o))]: ((((p@SV9)@SV4)=$false) | ((~ ((p@(^[SY27:mu,SY28:$i]: (~
((SV9@SY27)@SY28))))@SV4))=$true) | (((p@(^[SY27:mu,SY28:$i]: (~ ((SV9@SY27)@SY28))))@SV4)=
$false))),inference(extcnf_not_pos,[status(thm)],[100])).
 thf(105,plain,(![SV4:$i,SV9:(mu>($i>$o))]: ((((p@(^[SY27:mu,SY28:$i]: (~ ((SV9@SY27)@SY28))))@SV4)=
$false) | (((p@SV9)@SV4)=$false) | (((p@(^[SY27:mu,SY28:$i]: (~ ((SV9@SY27)@SY28))))@SV4)=
$false))),inference(extcnf_not_pos,[status(thm)],[103])).
 thf(107,plain,(![SV8:(mu>($i>$o)),SV3:$i,SV22:(mu>($i>$o))]:
((((SV22@(((sK2_SY33@SV3)@(^[SX0:mu,SX1:$i]: (~ ((SV22@SX0)@SX1))))@SV8))@(((sK1_SY31@(^[SX0:mu,SX1:$i]:
(~ ((SV22@SX0)@SX1))))@SV8)@SV3))=$true) | (((p@SV8)@SV3)=$false) | (((p@(^[SX0:mu,SX1:$i]: (~
((SV22@SX0)@SX1))))@SV3)=$true))),inference(extcnf_not_neg,[status(thm)],[78])).
 thf(108,plain,(![SV11:(mu>($i>$o)),SV3:$i,SV15:(mu>($i>$o))]:
((((SV15@(((sK2_SY33@SV3)@SV11)@(^[SX0:mu,SX1:$i]: (~
((SV15@SX0)@SX1)))))@(((sK1_SY31@SV11)@(^[SX0:mu,SX1:$i]: (~ ((SV15@SX0)@SX1))))@SV3))=$false) |
(((p@(^[SX0:mu,SX1:$i]: (~ ((SV15@SX0)@SX1))))@SV3)=$false) | (((p@SV11)@SV3)=
$true))),inference(extcnf_not_pos,[status(thm)],[81])).
 thf(109,plain,(![SV4:$i,SV9:(mu>($i>$o))]: ((((p@(^[SY27:mu,SY28:$i]: (~ ((SV9@SY27)@SY28))))@SV4)=
$false) | (((p@SV9)@SV4)=$false))),inference(sim,[status(thm)],[105])).
 thf(110,plain,(![SV4:$i,SV9:(mu>($i>$o))]: ((((p@SV9)@SV4)=$true) | (((p@(^[SY29:mu,SY30:$i]: (~
((SV9@SY29)@SY30))))@SV4)=$true))),inference(sim,[status(thm)],[101])).
 thf(111,plain,(![SV3:$i,SV8:(mu>($i>$o))]: ((((p@SV8)@SV3)=$false) | (((p@(^[SX0:mu,SX1:$i]:
$true))@SV3)=$true))),inference(sim,[status(thm)],[76])).
 thf(112,plain,(![SV11:(mu>($i>$o)),SV3:$i]: ((((p@(^[SX0:mu,SX1:$i]: $false))@SV3)=$false) |
(((p@SV11)@SV3)=$true))),inference(sim,[status(thm)],[80])).
 thf(113,plain,((($false)=$true)),inference(fo_atp_e,[status(thm)],
[25,112,111,110,109,108,107,84,83,82,75,74,73,72,71,70,69,68,67,66,65,62,57,56,51,42,29])).
 thf(114,plain,($false),inference(solved_all_splits,[solved_all_splits(join,[])],[113])).
% SZS output end CNFRefutation

%**** End of derivation protocol ****
%**** no. of clauses in derivation: 97 ****
%**** clause counter: 113 ****
```

Fig. 1. Excerpt of LEO-II's inconsistency proof (for Gödel's variant of the ontological argument).

theorem provers were available. In our computer-assisted study of the argument, the HOL ATPs LEO-II, Satallax and Nitpick have made some interesting observations [14]; the respective TPTP THF0 formalisation and further information is available online at http://github.com/FormalTheology/GoedelGod/.

In particular LEO-II was extensively used during the formalisation, and it was the first prover to fully automate the four steps as described in the notes on Gödel's proof by Dana Scott [35]. LEO-II's result was subsequently confirmed by Satallax. Interestingly, LEO-II can prove that Gödel's original axioms [30] are inconsistent: in these notes definition D2 (*An* essence *of an individual is a property possessed by it and necessarily implying any of its properties*: ϕ *ess.* $x \leftrightarrow \phi(x) \wedge \forall\psi[\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y))]$) is lacking conjunct $\phi(x)$, which has been added by Scott. Gödel's axioms are consistent only with this conjunct present. LEO-II's inconsistency result is new; it has not been reported in philosophy publications.

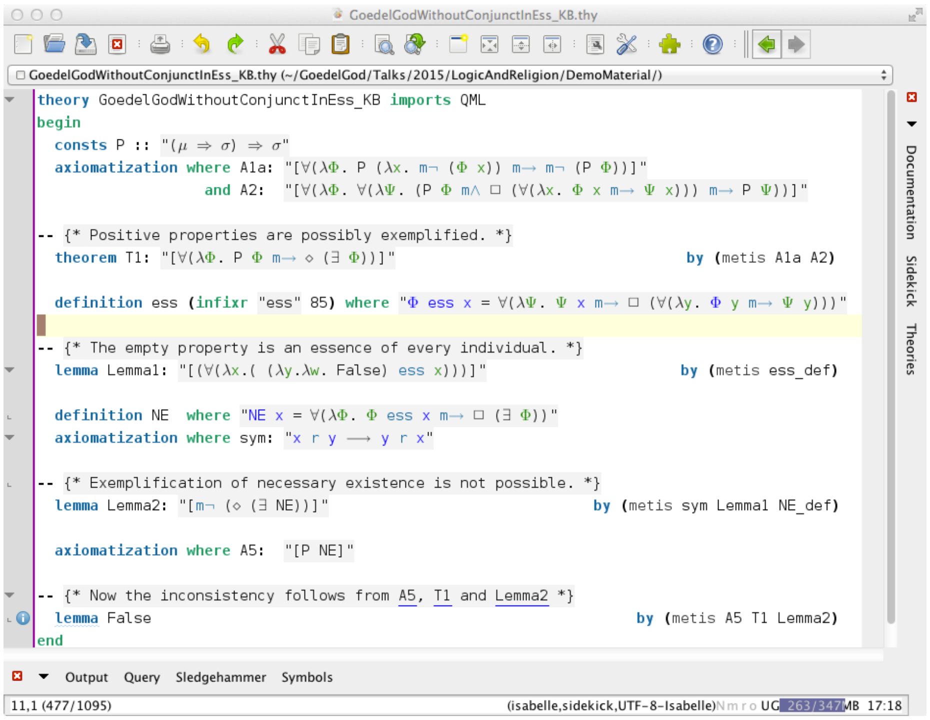


Fig. 2. Reconstruction and verification of LEO-II's inconsistency argument (for Gödel's variant of the ontological argument) in Isabelle.

Unfortunately, I have for a long time not been able to extract the key ideas of LEO-II's inconsistency proof. This has been due to a combination of aspects, including LEO-II's machine oriented (extensional) resolution calculus, the prover's human-unfriendly presentation of the generated proof object (cf. Fig. 1), and LEO's complex collaboration with external first-order ATPs, which could not easily be made fully transparent in the given case.

However, inspired by a discussion with Chad Brown on LEO-II's proof, we have recently been able to extract the core argument and reformulated and verified it as a human friendly, three step inconsistency argument in Isabelle. This reconstructed, intuitive argument can now even be automated with Metis; see Fig. 2. There are two core lemmata introduced, which, once they are revealed and experienced, appear very plausible ("the empty property is an essence of every individual" and "exemplification of necessary existence is not possible").

In the meantime, the HOL-ATPs have been successfully employed in further related experiments in metaphysics [21]. This includes the study and verification resp. falsification of follow-up papers on Gödel's work, which try to remedy a fundamental critique on the argument known as the modal collapse (this was brought up by Anderson [1]; the HOL ATPs reconfirmed it in our experiments):

both, Gödel's and Scott's formalisations, imply that $\forall\phi(\phi \rightarrow \Box\phi)$ holds, i.e. contigent truth implies necessary truth.

3 Summary

The embedding approach has many interesting advantages and it provides the probably most universal theorem proving approach to date that has actually been implemented and employed.

A key observation from our experiments in metaphysics is that the granularity levels of the philosophical arguments in the various papers we looked at is already well matched by today's automation capabilities of HOL ATPs. In nearly all cases the HOL ATPs either quickly confirmed the single argumentation steps or they presented a countermodel. This provides a good motivation for further application studies (not only) in metaphysics.

References

1. C. Anderson. Some emendations of Gödel's ontological proof. *Faith and Philosophy*, 7(3), 1990.
2. C. Benzmüller. Automating access control logic in simple type theory with LEO-II. In D. Gritzalis and J. López, editors, *SEC 2009*, volume 297 of *IFIP*, pages 387–398. Springer, 2009.
3. C. Benzmüller. Combining and automating classical and non-classical logics in classical higher-order logic. *Annals of Mathematics and Artificial Intelligence*, 62(1-2):103–128, 2011.
4. C. Benzmüller. Automating quantified conditional logics in HOL. In F. Rossi, editor, *IJCAI 2013*, pages 746–753, Beijing, China, 2013.
5. C. Benzmüller. Cut-free calculi for challenge logics in a lazy way. In C. N. Clint van Alten, Petr Cintula, editor, *Proceedings of the International Workshop on Algebraic Logic in Computer Science*, 2013.
6. C. Benzmüller. A top-down approach to combining logics. In *ICAART 2013*, pages 346–351, Barcelona, Spain, 2013. SciTePress Digital Library.
7. C. Benzmüller. Higher-order automated theorem provers. In D. Delahaye and B. Woltzenlogel Paleo, editors, *All about Proofs, Proof for All*, Mathematical Logic and Foundations, pages 171–214. College Publications, London, UK, 2015.
8. C. Benzmüller. HOL provers for first-order modal logics — experiments. In C. Benzmuüller and J. Otten, editors, *ARQNL@IJCAR 2014*, EPiC Series. EasyChair, 2015. To appear.
9. C. Benzmüller, M. Claus, and N. Sultana. Systematic verification of the modal logic cube in Isabelle/HOL. In C. Kaliszyk and A. Paskevich, editors, *PxTP 2015*, Berlin, Germany, 2015. EPTCS. To appear.
10. C. Benzmüller, D. Gabbay, V. Genovese, and D. Rispoli. Embedding and automating conditional logics in classical higher-order logic. *Annals of Mathematics and Artificial Intelligence*, 66(1-4):257–271, 2012.
11. C. Benzmüller and D. Miller. Automation of higher-order logic. In D. M. Gabbay, J. H. Siekmann, , and J. Woods, editors, *Handbook of the History of Logic, Volume 9 — Computational Logic*, pages 215–254. North Holland, Elsevier, 2014.

12. C. Benzmüller, J. Otten, and T. Raths. Implementing and evaluating provers for first-order modal logics. In L. D. Raedt, C. Bessiere, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, and P. Lucas, editors, *ECAI 2012*, volume 242 of *Frontiers in Artificial Intelligence and Applications*, pages 163–168, Montpellier, France, 2012. IOS Press.
13. C. Benzmüller and B. W. Paleo. Gödel's God in Isabelle/HOL. *Archive of Formal Proofs*, 2013, 2013.
14. C. Benzmüller and B. W. Paleo. Automating Gödel's ontological proof of God's existence with higher-order automated theorem provers. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *ECAI 2014*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 93 – 98. IOS Press, 2014.
15. C. Benzmüller and L. Paulson. Exploring properties of normal multimodal logics in simple type theory with LEO-II. In C. Benzmüller, C. Brown, J. Siekmann, and R. Statman, editors, *Reasoning in Simple Type Theory — Festschrift in Honor of Peter B. Andrews on His 70th Birthday*, Studies in Logic, Mathematical Logic and Foundations, pages 386–406. College Publications, 2008.
16. C. Benzmüller and L. Paulson. Multimodal and intuitionistic logics in simple type theory. *The Logic Journal of the IGPL*, 18(6):881–892, 2010.
17. C. Benzmüller and L. Paulson. Quantified multimodal logics in simple type theory. *Logica Universalis (Special Issue on Multimodal Logics)*, 7(1):7–20, 2013.
18. C. Benzmüller, L. C. Paulson, N. Sultana, and F. Theiß. The higher-order prover LEO-II. *Journal of Automated Reasoning*, 2015. To appear.
19. C. Benzmüller and A. Pease. Higher-order aspects and context in SUMO. *Journal of Web Semantics (Special Issue on Reasoning with context in the Semantic Web)*, 12-13:104–117, 2012.
20. C. Benzmüller and T. Raths. HOL based first-order modal logic provers. In K. L. McMillan, A. Middeldorp, and A. Voronkov, editors, *LPAR 2013*, volume 8312 of *LNCS*, pages 127–136, Stellenbosch, South Africa, 2013. Springer.
21. C. Benzmüller, L. Weber, and B. Woltzenlogel Paleo. Computer-assisted analysis of the Anderson-Hájek ontological controversy. In R. S. Silvestre and J.-Y. Béziau, editors, *Handbook of the 1st World Congress on Logic and Religion, Joao Pessoa, Brasil*, pages 53–54, 2015.
22. C. Benzmüller and B. Woltzenlogel Paleo. Higher-order modal logics: Automation and applications. In A. Paschke and W. Faber, editors, *Reasoning Web 2015*, number 9203 in LNCS, pages 1–43, Berlin, Germany, 2015. Springer. To appear.
23. C. Benzmüller and B. Woltzenlogel Paleo. Interacting with modal logics in the coq proof assistant. In *CSR 2015*, volume 9139 of *LNCS*, pages 398–411. Springer, 2015. To appear.
24. Y. Bertot and P. Casteran. *Interactive Theorem Proving and Program Development*. Springer, 2004.
25. J. Blanchette, S. Böhme, and L. Paulson. Extending Sledgehammer with SMT solvers. *Journal of Automated Reasoning*, 51(1):109–128, 2013.
26. J. C. Blanchette and T. Nipkow. Nitpick: A counterexample generator for higher-order logic based on a relational model finder. In M. Kaufmann and L. C. Paulson, editors, *ITP 2010*, volume 6172 of *LNCS*, pages 131–146. Springer, 2010.
27. C. E. Brown. Satallax: An automated higher-order prover. In B. Gramlich, D. Miller, and U. Sattler, editors, *JCAR 2012*, volume 7364 of *LNCS*, pages 111 – 117. Springer, 2012.
28. A. Bundy. The use of explicit plans to guide inductive proofs. In E. L. Lusk and R. A. Overbeek, editors, *CADE 1988*, volume 310 of *LNCS*, pages 111–120. Springer, 1988.

29. D. M. Gabbay. *Labelled Deductive Systems*. Clarendon Press, 1996.
30. K. Gödel. *Appx.A: Notes in Kurt Gödel's Hand*, pages 144–145. In [37], 2004.
31. E. Melis, A. Meier, and J. H. Siekmann. Proof planning with multiple strategies. *Artif. Intell.*, 172(6-7):656–684, 2008.
32. T. Nipkow, L. C. Paulson, and M. Wenzel. *Isabelle/HOL - A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, 2002.
33. H. J. Ohlbach. Semantics-based translation methods for modal logics. *Journal of Logic and Computation*, 1(5):691–746, 1991.
34. H. J. Ohlbach and R. A. Schmidt. Functional translation and second-order frame properties of modal logics. *Journal of Logic and Computation*, 7(5):581–603, 1997.
35. D. Scott. *Appx.B: Notes in Dana Scott's Hand*, pages 145–146. In [37], 2004.
36. J. H. Siekmann, C. Benzmüller, and S. Autexier. Computer supported mathematics with omega. *J. Applied Logic*, 4(4):533–559, 2006.
37. J. Sobel. *Logic and Theism: Arguments for and Against Beliefs in God*. Cambridge U. Press, 2004.
38. G. Sutcliffe. The TPTP problem library and associated infrastructure. *Journal of Automated Reasoning*, 43(4):337–362, 2009.
39. G. Sutcliffe and C. Benzmüller. Automated reasoning in higher-order logic using the TPTP THF infrastructure. *Journal of Formalized Reasoning*, 3(1):1–27, 2010.
40. M. Wisnieski and A. Steen. Embedding of quantified higher-order nominal modal logic into classical higher-order logic. In C. Benzmuüller and J. Otten, editors, *Proceedings on the 1st International Workshop on Automated Reasoning in Quantified Non-Classical Logics (ARQNL)*, 2014.