Automated Reasoning in Modal Logics: A Framework with Applications

Branden Fitelson

Department of Philosophy
Institute for Cognitive and Brain Sciences
Group in Logic and Methodology of Science

University of California–Berkeley

branden@fitelson.org
http://fitelson.org/

August 2005

The principle that every truth is possibly necessary can now be shown to entail that every truth is necessary by a chain of elementary inferences in a perspicuous notation unavailable to Hegel. —Williamson [5, p. 4]

Here's what Williamson means by "perspicuous notation":

- ⌜$\Box p$⌝ for ⌜It is necessarily the case that p⌝.
- ⌜$\Diamond p$⌝ for ⌜It is possibly the case that p⌝.
- ⌜$p \rightarrow q$⌝ for ⌜p implies q⌝.
- ⌜$\sim p$⌝ for ⌜It is not the case that p⌝.

…in the context of **K**, *the adoption of* [**F**] *is equivalent to the adoption of* $\mathbf{M}_{\Box}$, *the quasi-Megarian axiom. … Yet just before … Aristotle had … argued against the Megarian position. And so how can he now … be propounding a principle that commits him to a position that he had previously rejected?* —Fine [2, p. 8]

W&F are working with the classical $\langle \rightarrow, \sim \rangle$-sentential modal logics **CK** and **CKT**, which can be characterized Hilbert-Style [6]:

Modus Ponens Rule Schema (MP)

- If $\vdash p$ and $\vdash p \rightarrow q$, then $\vdash q$ [either $\nvdash p$ or $\nvdash p \rightarrow q$ or $\vdash q$].

(Logical) (MP)–Axiom Schemata for $\langle \rightarrow, \sim \rangle$-Classical Logic (**C**) [3]

- ($\mathbf{C}_1$) $\vdash (p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r))$
- ($\mathbf{C}_2$) $\vdash p \rightarrow (\sim p \rightarrow q)$
- ($\mathbf{C}_3$) $\vdash (\sim p \rightarrow p) \rightarrow p$

Necessitation Rule Schema (RN):

- If $\vdash p$, then $\vdash \Box p$ [either $\nvdash p$ or $\vdash \Box p$].

(Proper) Modal Axiom Schemata:

- (**K**) $\vdash \Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$
- (**T**) $\vdash \Box p \rightarrow p$
- (Def.$_\Diamond$) $\Diamond p \stackrel{\text{def}}{=} \sim\Box\sim p$ [$\Diamond p$ and $\sim\Box\sim p$ are intersubstitutable]

OK. That's the logical background. Now, we're ready to go . . .

- *Williamson's [5] triviality result*: If (**W**) $p \to \Diamond\Box p$ is added as an axiom schema to **CKT**, then (**M**) $p \leftrightarrow \Box p$ is a theorem (both $p \to \Box p$ and $\Box p \to p$ are theorem schemata of **CKTW**).
- *Two of Fine's [2] triviality results*. Both involve the schema:

(**F**) $$\Box(\Diamond p \to \Diamond q) \to \Box(p \to q)$$

  - (**M**) $p \leftrightarrow \Box p$ is a theorem schema of the system **CKTF**.
  - ($\mathbf{M}_{\Box}$) $\Box(p \leftrightarrow \Diamond p)$ is a theorem schema of the system **CKF**.
- Proving these triviality results is (increasingly) non-trivial!
- In the remainder of this talk, I will do three things:
  - Explain how to get `Otter` to prove these triviality results.
  - Explain how to get `Otter` to prove more general t-results.
  - Explain how to get `Paradox` to prove **non**-triviality results.
- The key step is *representing* enough of the metatheory of $\langle\to, \sim\rangle$-sentential modal logics in simple, first-order terms.
- The rest is just (only moderately skilled) application of the first-order order AR programs `Otter` [4] and `Paradox` [1].

- Happily, we can simply express (enough of) the metatheory of the salient $\langle \rightarrow, \sim \rangle$-sentential modal logics in FOL:
  - Monadic predicate 'P': interpreted as 'is a theorem' ($\vdash$).
  - Logical Operator '-': metalanguage negation sign.
  - Logical Operator '|': metalanguage disjunction sign.
  - Unary function 'n': object language negation operator ($\sim$).
  - Unary function 'l': object language necessity operator ($\Box$).
  - Binary function 'i': object lang. implication operator ($\rightarrow$).
- Here are all of our rule and axiom schemata, expressed in our FOL as *clauses* (*i.e.*, in implicitly $\forall$-quantified CNF):
  - (MP): -P(x) | -P(i(x,y)) | P(y).
  - (**C**$_1$): P(i(i(x,y),i(i(y,z),i(x,z)))).
  - (**C**$_2$): P(i(x,i(n(x),y))).
  - (**C**$_3$): P(i(i(n(x),x),x)).
  - (RN): -P(x) | P(l(x)).
  - (**K**): P(i(l(i(x,y)),i(l(x),l(y)))).
  - (**T**): P(i(l(x),x)).
  - (**W**): P(i(x,n(l(n(l(x)))))).
  - (**F**): P(i(l(i(n(l(n(x))),n(l(n(y))))),l(i(x,y)))).

- Now that we have everything we need expressed in clausal form, we can simply feed in various problems to `Otter` [4] and `Paradox` [1], and see what happens. Stuff happens . . .
- `Otter` proves both Williamson's and Fine's triviality results without too much difficulty (using sufficient knowledge about solving these sorts of problems with `Otter`!).
- But, that's just the beginning! The real power of `Otter` is in its ability to *generalize* these triviality results. Most impressive are generalizations to *non-classical logics*.
- I will focus on Fine's second triviality result. Recall, this result is that $\Box(p \leftrightarrow \Diamond p)$ is a theorem schema of the system **CKF**. Question: What happens when we *weaken* **C** here?
- Fine's proof of this result is strongly classical in nature. So, one might suspect that the result does not generalize to underlying logics weaker than **C**. This is far from the truth.
- In fact, for a wide variety of logics **X**, the modal logic **XKF** has $\Box(p \leftrightarrow \Diamond p)$ as a theorem schema. More precisely . . .

- Using Otter, I found proofs of $\Box(p \leftrightarrow \Diamond p)$ in intuitionistic, three-valued (both Kleene and Łukasiewicz 3-valued), and infinite-valued modal-logics. Then, I did something strange.
- I took the *intersection* of all of these non-classical Otter proofs, and I discovered that the following four underlying $\langle\rightarrow, \sim\rangle$-schemata are sufficient to generate Fine's triviality:
  - $\vdash (p \rightarrow q) \rightarrow ((r \rightarrow p) \rightarrow (r \rightarrow q))$
  - $\vdash (p \rightarrow q) \rightarrow (\sim q \rightarrow \sim p)$
  - $\vdash (p \rightarrow (q \rightarrow r)) \rightarrow (q \rightarrow (p \rightarrow r))$
  - $\vdash ((p \rightarrow q) \rightarrow r) \rightarrow (q \rightarrow r)$
- Moreover, these four schema (plus MP) are *independent*. So, this is a very weak *logical basis* for Fine's triviality result.
- The proofs of this generalization of Fine's triviality result are highly non-trivial! Note: there is no Kripke semantics for this modal system, so one is *forced* to work axiomatically!
- Q: if there are no Kripke semantics for these kinds of weak, non-classical modal systems, then how could we ever prove that a modal logic of this kind is **non**-trivial? A: Paradox!

- `Paradox` [1] is a recent program, which finds (relatively small) models for (finitely satisfiable!) sets of FOL clauses.
- `Paradox` is an order of magnitude more efficient than previous FO model-finders, for problems of the kind we are discussing. It can find logics with up to 16 values [$\gg 5$].
- For instance, Williamson notes that *his* triviality disappears if we replace our *classical* underlying $\langle \rightarrow, \sim \rangle$-sentential logic **C** with the *intuitionistic* underlying $\langle \rightarrow, \sim \rangle$-sentential logic **H**.
- `Paradox` easily finds a 4-valued logic in which (MP) and (RN) preserve theoremhood; the axioms of (**H**) are all theorems; (**K**), (**T**), and (**W**) are all theorems; but $\nvdash p \rightarrow \Box p$.
- `Paradox` also allows us to show that various (even weaker) non-classical logics are *too weak* to generate Fine's triviality.
- *E.g.*, `Paradox` allows us to prove (*via* a 4-valued logic) that the relevance logic **E** is *too weak* to generate Fine's paradox [**EKW** $\nvdash \Box(p \leftrightarrow \Diamond p)$]. Open Question: **RKW** $\vdash \Box(p \leftrightarrow \Diamond p)$?

- I showed how we can express the metatheory of (almost all) sentential modal logics in elementary, FOL terms.
- This allows us to use first-order theorem-provers like `Otter` to prove interesting and non-trivial theorems in just about any sentential modal logic you can cook up.
- And, we can use 1st-order model-finders like `Paradox` to establish **non**-theoremhood in just about any sentential modal logic you can cook up (*no Kripke semantics needed!*).
- I used, as illustrations, both positive and negative results concerning the triviality of some modal systems from recent philosophical discussions of Fine [2] & Williamson [5].
- I have many additional results (both positive and negative) concerning various other non-classical underlying logics, as well as combinations involving various other modal axioms.
- The status of only a few of the plethora of resulting "triviality questions" remains open. This is a testament to the power of `Otter` & `Paradox` in solving such problems.

[1] K. Claessen and N. Sörensson, “New Techniques that Improve MACE-style Finite Model Finding”, 2003, manuscript. URL: www.cs.chalmers.se/~koen/Papers/paradox.ps.

[2] K. Fine, “Aristotle’s Megarian Maneuvers”, 2004, manuscript, which can be downloaded from Fine’s website: philosophy.fas.nyu.edu/docs/IO/1160/aristotlemodalitydraft.pdf.

[3] J. Łukasiewicz, *Elements of Mathematical Logic*, 1964 [translation of his 1929 *Elementy logiki matematycznej*].

[4] W. McCune, *Otter 3.3 Reference Manual*. Tech. Memo ANL/MCS-TM-263, MCS-ANL, 2003, which can be downloaded from: www-unix.mcs.anl.gov/AR/otter/otter33.pdf.

[5] T. Williamson, “Must Do Better”, 2005, manuscript, which can be downloaded from Williamson’s website: www.philosophy.ox.ac.uk/faculty/members/docs/Must%20Do%20Better.pdf.

[6] J. Zeman, *Modal Logic*, 1973. This classic text can now be downloaded from: www.clas.ufl.edu/users/jzeman/modallogic/.