Limbo

A **Reasoning System** for a **First-Order** Logic of **Limited Belief**

Christoph Schwering

UNSW Sydney

What is limited belief? And why?

Task: Robot has a KB and a query:

Does the KB *logically entail* the query **?**

What is limited belief? And why?

Task: Robot has a KB and a query:

Does the KB $\underbrace{\textit{logically entail}}_{\text{Which logic?}}$ the query **?**

What is limited belief? And why?

Task: Robot has a KB and a query:

Does the KB $\underbrace{\textit{logically entail}}_{\text{Which logic?}}$ the query **?**

Classical logic:

- Unrealistic: omniscient agent
- Undecidable (first-order) / intractable (propositional)

What is limited belief? And why?

Task: Robot has a KB and a query:

Does the KB $\underbrace{\textit{logically entail}}_{\text{Which logic?}}$ the query **?**

Limited belief:

- Belief level 0: explicitly written down in the KB
- Belief level $k > 0$: derivable from KB with effort k

Hope: good results at *small* belief level

Builds on Lakemeyer & Levesque, KR-2016

Language

FOL with equality + functions + sorts +

- Knowledge: $\mathbf{K}_0\,\alpha \quad \mathbf{K}_1\,\alpha \quad \mathbf{K}_2\,\alpha \quad \ldots$
- Possibility: $\mathbf{M}_0\,\alpha \quad \mathbf{M}_1\,\alpha \quad \mathbf{M}_2\,\alpha \quad \ldots$

Example:

- $\mathbf{K}_1\,\big(\mathrm{Rich}(\mathrm{Frank}) \vee \mathrm{Rich}(\mathrm{Fred})\big)$ — know that Frank or Fred is rich
- $\forall x\,\mathbf{M}_1\,\mathrm{fatherOf}(\mathrm{Sally}) \neq x$ — don't know who Sally's father is
- $\mathbf{K}_1\,\exists x\,\big(\mathrm{fatherOf}(\mathrm{Sally}) = x \wedge \mathrm{Rich}(x) \wedge \mathbf{M}_1\,\mathrm{fatherOf}(\mathrm{Sally}) \neq x\big)$ — know that Sally's father is rich, but don't know who he is

Semantics

Model: set of **clauses** closed under unit propagation

- Belief level 0: subsumption
- Belief level $k > 0$: k case splits

Example:

If all we know is (a) $\mathrm{fatherOf}(\mathrm{Sally}) = \mathrm{Frank} \lor \mathrm{fatherOf}(\mathrm{Sally}) = \mathrm{Fred}$

and (b) $\forall x \, (\mathrm{fatherOf}(\mathrm{Sally}) \neq x \lor \mathrm{Rich}(x))$

then $\mathbf{K}_1 \, (\mathrm{Rich}(\mathrm{Frank}) \lor \mathrm{Rich}(\mathrm{Fred}))$ **?**

Yes! Branch on $\mathrm{fatherOf}(\mathrm{Sally})$:

- $\{$(a), (b), $\mathrm{fatherOf}(\mathrm{Sally}) = \mathrm{Frank}\} \ni \mathrm{Rich}(\mathrm{Frank})$ by UP with (b)
- $\{$(a), (b), $\mathrm{fatherOf}(\mathrm{Sally}) = \mathrm{Fred}\} \ni \mathrm{Rich}(\mathrm{Fred})$ by UP with (b)
- $\{$(a), (b), $\mathrm{fatherOf}(\mathrm{Sally}) = n\} \ni \bot$ by UP with (a)
  for $n \neq \mathrm{Frank}, \mathrm{Fred}$

$\mathrm{Rich}(\mathrm{Frank}) \lor \mathrm{Rich}(\mathrm{Fred})$

KB entails query at some belief level $\Longrightarrow$ KB classically entails query

if no $\neg\mathbf{K}$, $\neg\mathbf{M}$

KB entails query at some belief level $\Longleftrightarrow$ KB classically entails query
if no $\neg\mathbf{K}$, $\neg\mathbf{M}$ and no $\exists, \forall$

KB entails query at some belief level is decidable

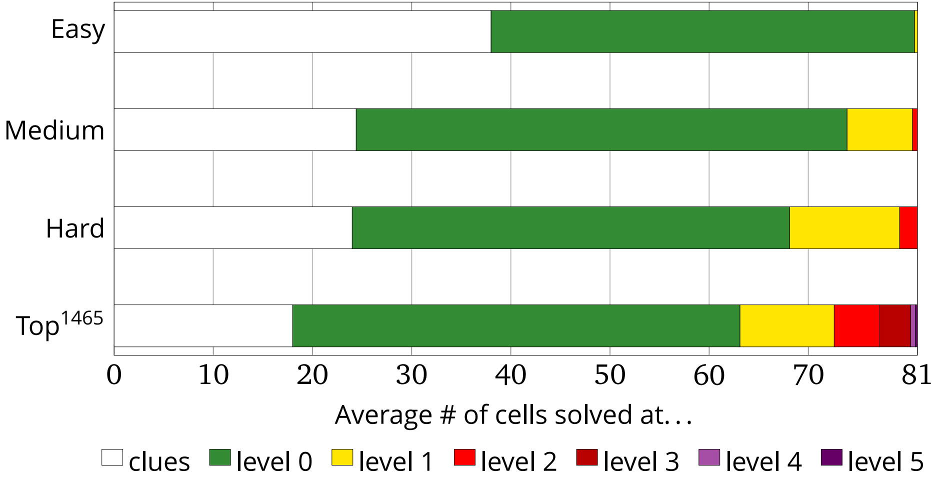
KB entails query at some belief level is tractable
if no $\exists, \forall$ and belief level fixed

Experiments: Sudoku Minesweeper

Hypothesis: good results at *small* belief level

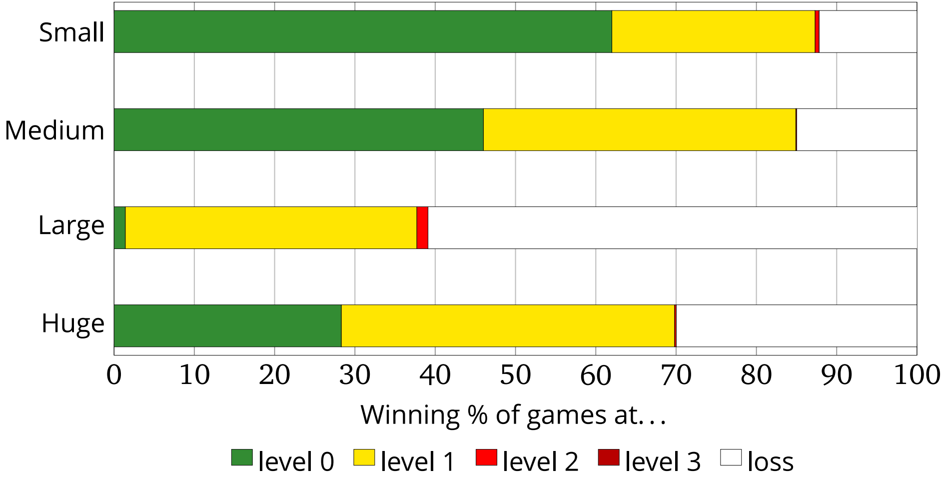
Experiments: Sudoku Minesweeper

<u>Hypothesis</u>: good results at *small* belief level ✓



Experiments: Sudoku Minesweeper

Hypothesis: good results at *small* belief level ✓ ✓



Limbo = Limited Belief

Demos: `www.cse.unsw.edu.au/~cschwering/limbo`
Fri 10:00–12:00

Code: `www.github.com/schwering/limbo`

Next: 1. actions 2. belief change 3. multiple agents

Appendix

Language in detail

Terms:

- First-order variables
- Functions $f(t_1, \ldots, t_m)$ where each t_i is a name or variable
- Standard names infinitely many and sorted

Formulas:

- FOL: $t_1 = t_2 \quad \neg\alpha \quad \alpha \vee \beta \quad \exists x\, \alpha$
- Knowledge: $\mathbf{K}_0\, \alpha \quad \mathbf{K}_1\, \alpha \quad \mathbf{K}_2\, \alpha \quad \ldots$
- Possibility: $\mathbf{M}_0\, \alpha \quad \mathbf{M}_1\, \alpha \quad \mathbf{M}_2\, \alpha \quad \ldots$
- Knowledge base: $\mathbf{O}\, \alpha$ where α is in universal CNF

- $\alpha \wedge \beta \quad \alpha \supset \beta \quad \alpha \equiv \beta \quad \forall x\, \alpha$ are abbreviations
- Predicates are simulated with functions
- Existentials in KBs are simulated with Skolem functions
- Functions on the right-hand side and within functions are flattened:

$$f(\cdot) = g(\cdot) \quad \mapsto \quad \forall x\, (g(\cdot) = x \supset f(\cdot) = x)$$
$$f(g(\cdot)) = t \quad \mapsto \quad \forall x\, (g(\cdot) = x \supset f(x) = t)$$

Literal encoding

- Functions cannot appear on rhs $\quad f(\cdot) = g(\cdot) \quad \mapsto \quad \forall x\,(g(\cdot) = x \supset f(\cdot) = x)$
- Functions cannot be nested $\quad f(g(\cdot)) = t \quad \mapsto \quad \forall x\,(g(\cdot) = x \supset f(x) = t)$

- Term is 30-bit number
  - points to full representation
  - this pointer is unique (interning)
- Literal is 64-bit number
  - $30 + 30$ bits for lhs + rhs
  - $1 + 1$ bits to indicate if lhs + rhs is name
  - 1 bit to indicate whether $=$ or $\neq$

- Conditions for literal <u>subsumption</u> and <u>complementarity</u>:
  - ℓ subsumes ℓ
  - $t = n_1$ subsumes $t \neq n_2$
  - $t = t'$ and $t \neq t'$ are complementary
  - $t = n_1$ and $t = n_2$ are complementary

  t, t' ground terms
  n_1, n_2 distinct names
- Sound and complete
- Bitwise op's on 64-bit numbers suffice no term dereferencing
- Fast clause subsumption and unit propagation

"I don't know Sally's father, but I know he's rich"

- $c_1 = \mathrm{f}(\mathrm{S}) = \mathrm{Frank} \vee \mathrm{f}(\mathrm{S}) = \mathrm{Fred}$
  $c_2 = \forall x\,(\mathrm{f}(\mathrm{S}) \neq x \vee \mathrm{r}(x) = \top)$
- $\mathbf{O}(c_1 \wedge c_2) \models \mathbf{K}\exists x \left(\mathrm{f}(\mathrm{S}) = x \wedge \mathrm{r}(x) = \top \wedge \mathbf{M}\,\mathrm{f}(\mathrm{S}) \neq x\right)$

"I don't know Sally's father, but I know he's rich"

- $e = \{w \mid w \models \mathrm{f}(\mathrm{S}) = \mathrm{Frank} \vee \mathrm{f}(\mathrm{S}) = \mathrm{Fred} \wedge$
  $\forall x\, (\mathrm{f}(\mathrm{S}) \neq x \vee \mathrm{r}(x) = \top)\}$
- $e \models \mathbf{K} \exists x \left(\mathrm{f}(\mathrm{S}) = x \wedge \mathrm{r}(x) = \top \wedge \mathbf{M}\, \mathrm{f}(\mathrm{S}) \neq x \right)$

“I don’t know Sally’s father, but I know he’s rich”

- $e = \{w \mid w \models \mathrm{f}(\mathrm{S}) = \mathrm{Frank} \vee \mathrm{f}(\mathrm{S}) = \mathrm{Fred} \wedge$
  $\forall x\,(\mathrm{f}(\mathrm{S}) \neq x \vee \mathrm{r}(x) = \top)\}$
- $e \models \mathbf{K}\exists x \left(\mathrm{f}(\mathrm{S}) = x \wedge \mathrm{r}(x) = \top \wedge \mathbf{M}\,\mathrm{f}(\mathrm{S}) \neq x\right)$

- For every $w \in e$, for some n, $w \models \mathrm{f}(\mathrm{S}) = n \wedge \mathrm{R}(n)$
- For some $w' \in e$, $w \models \mathrm{f}(\mathrm{S}) \neq n$

"I don't know Sally's father, but I know he's rich"

- $c_1 = \mathrm{f}(\mathrm{S}) = \mathrm{Frank} \vee \mathrm{f}(\mathrm{S}) = \mathrm{Fred}$
  $c_2 = \forall x\,(\mathrm{f}(\mathrm{S}) \neq x \vee \mathrm{r}(x) = \top)$
- $\mathbf{O}(c_1 \wedge c_2) \mathrel{|\!\!\approx} \mathbf{K}_1 \exists x \left(\mathrm{f}(\mathrm{S}) = x \wedge \mathrm{r}(x) = \top \wedge \mathbf{M}_1 \mathrm{f}(\mathrm{S}) \neq x\right)$

“I don’t know Sally’s father, but I know he’s rich”

- $s = \{\mathrm{f}(\mathrm{S}) = \mathrm{Frank} \lor \mathrm{f}(\mathrm{S}) = \mathrm{Fred},$
  $\mathrm{f}(\mathrm{S}) \neq n \lor \mathrm{r}(n) = \top \mid n$ is a name$\}$
- $s \mathrel{|\!\!\approx} \mathbf{K}_1 \exists x \left(\mathrm{f}(\mathrm{S}) = x \land \mathrm{r}(x) = \top \land \mathbf{M}_1 \mathrm{f}(\mathrm{S}) \neq x\right)$

"I don't know Sally's father, but I know he's rich"

- $s = \{\mathrm{f}(\mathrm{S}) = \mathrm{Frank} \vee \mathrm{f}(\mathrm{S}) = \mathrm{Fred},$
  $\mathrm{f}(\mathrm{S}) \neq n \vee \mathrm{r}(n) = \top \mid n$ is a name$\}$
- $s \approx\!\!\!| \ \mathbf{K}_1 \exists x \big(\mathrm{f}(\mathrm{S}) = x \wedge \mathrm{r}(x) = \top \wedge \mathbf{M}_1 \mathrm{f}(\mathrm{S}) \neq x\big)$

$\iff$

for some t_1, for all n_1, for some n,
$s \cup \{t_1 = n_1\} \approx\!\!\!| \ \mathrm{f}(\mathrm{S}) = n \wedge \mathrm{r}(n) = \top \wedge \mathbf{M}_1 \mathrm{f}(\mathrm{S}) \neq n$

“I don’t know Sally’s father, but I know he’s rich”

- $s = \{\mathrm{f}(\mathrm{S}) = \text{Frank} \vee \mathrm{f}(\mathrm{S}) = \text{Fred},$
  $\mathrm{f}(\mathrm{S}) \neq n \vee \mathrm{r}(n) = \top \mid n$ is a name$\}$
- $s \approx\!\!\!\mid\, \mathbf{K}_1 \exists x \left(\mathrm{f}(\mathrm{S}) = x \wedge \mathrm{r}(x) = \top \wedge \mathbf{M}_1 \mathrm{f}(\mathrm{S}) \neq x\right)$

$\iff$

for some t_1, for all n_1, for some n,
$s \cup \{t_1 = n_1\} \approx\!\!\!\mid\, \mathrm{f}(\mathrm{S}) = n \wedge \mathrm{r}(n) = \top \wedge \mathbf{M}_1 \mathrm{f}(\mathrm{S}) \neq n$

$\iff$

for some t_1, for all n_1, for some n, $s \cup \{t_1 = n_1\} \approx\!\!\!\mid\, \mathrm{f}(\mathrm{S}) = n \wedge \mathrm{r}(n) = \top$
for some t_2 and n_2, $s \cup \{t_2 = n_2\} \approx\!\!\!\mid\, \mathrm{f}(\mathrm{S}) \neq n$

"I don't know Sally's father, but I know he's rich"

- $s = \{\mathrm{f}(\mathrm{S}) = \mathrm{Frank} \vee \mathrm{f}(\mathrm{S}) = \mathrm{Fred},$
  $\mathrm{f}(\mathrm{S}) \neq n \vee \mathrm{r}(n) = \top \mid n$ is a name$\}$
- $s \mathrel{|\!\approx} \mathbf{K}_1 \exists x \left(\mathrm{f}(\mathrm{S}) = x \wedge \mathrm{r}(x) = \top \wedge \mathbf{M}_1 \mathrm{f}(\mathrm{S}) \neq x\right)$

$\iff$

(a) for some t_1, for all n_1, for some n, $s \cup \{t_1 = n_1\} \mathrel{|\!\approx} \mathrm{f}(\mathrm{S}) = n \wedge \mathrm{r}(n) = \top$
(b) for some t_2 and n_2, $s \cup \{t_2 = n_2\} \mathrel{|\!\approx} \mathrm{f}(\mathrm{S}) \neq n$

$\Longleftarrow$

(a) choose $t_1 = \mathrm{f}(\mathrm{S})$:
 if $n_1 = \mathrm{Frank}$, choose $n = \mathrm{Frank}$:
  $s \cup \{\mathrm{f}(\mathrm{S}) = \mathrm{Frank}\}$ contains $\mathrm{f}(\mathrm{S}) = \mathrm{Frank}, \mathrm{r}(\mathrm{Frank}) = \top$
(b) choose $t_2 = \mathrm{f}(\mathrm{S})$ and $n_2 = \mathrm{Fred}$:
   $s \cup \{\mathrm{f}(\mathrm{S}) = \mathrm{Fred}\}$ contains $\mathrm{f}(\mathrm{S}) \neq \mathrm{Frank}$
 if $n_1 = \mathrm{Fred}$: analogous
 if $n_1 \neq \mathrm{Frank}, \mathrm{Fred}$: $s \cup \{\mathrm{f}(\mathrm{S}) = n_1\}$ is obv. inconsistent

Theorems in detail

- $\models$ is classical entailment
- $\mathrel{|\!\approx}$ is limited entailment
- σ contains no $\mathbf{O}, \neg\mathbf{K}_k, \neg\mathbf{M}_k$
- $\sigma^\star$ removes belief levels
- σ_k sets belief levels to k

Soundness & Eventual Completeness

$\mathbf{O}\alpha \mathrel{|\!\approx} \sigma \quad \Longrightarrow \quad \mathbf{O}\alpha \models \sigma^\star$ if σ without $\neg\mathbf{K}_k, \neg\mathbf{M}_k$

$\mathbf{O}\alpha \mathrel{|\!\approx} \sigma_k$ for some $k \quad \Longleftrightarrow \quad \mathbf{O}\alpha \models \sigma^\star$ if α, σ quantifier-free

Complexity

$\mathbf{O}\alpha \mathrel{|\!\approx} \sigma$ is decidable

$\mathbf{O}\alpha \mathrel{|\!\approx} \sigma_k$ is tractable in $\mathcal{O}(2^k(|\alpha| + |\sigma|)^{k+3})$ if α, σ quantifier-free

Semantics in detail

- $(\neg)t = n$
- $(\alpha \vee \beta)$
- $\neg(\alpha \vee \beta)$
- $\exists x\, \alpha$
- $\neg\exists x\, \alpha$
- $\neg\neg\alpha$

- $\mathbf{K}_0\, \alpha$
- $\mathbf{K}_{k+1}\, \alpha$
- $\mathbf{M}_0\, \alpha$
- $\mathbf{M}_{k+1}\, \alpha$
- $\mathbf{O}\, \alpha$

Semantics in detail

- $s \approx\!\!\!| \; (\neg)t = n$ *iff* $(\neg)t = n \in s$
- $s \approx\!\!\!| \; (\alpha \vee \beta)$ *iff* $(\alpha \vee \beta) \in s$ or $s \approx\!\!\!| \; \alpha$ or $s \approx\!\!\!| \; \beta$
- $s \approx\!\!\!| \; \neg(\alpha \vee \beta)$ *iff* $s \approx\!\!\!| \; \neg\alpha$ and $s \approx\!\!\!| \; \neg\beta$
- $s \approx\!\!\!| \; \exists x\, \alpha$ *iff* $s \approx\!\!\!| \; \alpha^x_n$ for some name n
- $s \approx\!\!\!| \; \neg\exists x\, \alpha$ *iff* $s \approx\!\!\!| \; \neg\alpha^x_n$ for every name n
- $s \approx\!\!\!| \; \neg\neg\alpha$ *iff* $s \approx\!\!\!| \; \alpha$

- $\mathbf{K}_0\, \alpha$
- $\mathbf{K}_{k+1}\, \alpha$
- $\mathbf{M}_0\, \alpha$
- $\mathbf{M}_{k+1}\, \alpha$
- $\mathbf{O}\, \alpha$

Semantics in detail

- $s \mathrel{|\!\approx} (\neg)t = n$ *iff* $(\neg)t = n \in s$
- $s \mathrel{|\!\approx} (\alpha \vee \beta)$ *iff* $(\alpha \vee \beta) \in s$ or $s \mathrel{|\!\approx} \alpha$ or $s \mathrel{|\!\approx} \beta$
- $s \mathrel{|\!\approx} \neg(\alpha \vee \beta)$ *iff* $s \mathrel{|\!\approx} \neg\alpha$ and $s \mathrel{|\!\approx} \neg\beta$
- $s \mathrel{|\!\approx} \exists x\, \alpha$ *iff* $s \mathrel{|\!\approx} \alpha^x_n$ for some name n
- $s \mathrel{|\!\approx} \neg\exists x\, \alpha$ *iff* $s \mathrel{|\!\approx} \neg\alpha^x_n$ for every name n
- $s \mathrel{|\!\approx} \neg\neg\alpha$ *iff* $s \mathrel{|\!\approx} \alpha$

- $s \mathrel{|\!\approx} \mathbf{K}_0\, \alpha$ *iff* s is obviously inconsistent or $s \mathrel{|\!\approx} \alpha$
- $s \mathrel{|\!\approx} \mathbf{K}_{k+1}\, \alpha$ *iff* for some t and all n, $s \cup \{t = n\} \mathrel{|\!\approx} \mathbf{K}_k\, \alpha$
- $s \mathrel{|\!\approx} \mathbf{M}_0\, \alpha$ *iff* s is obviously consistent and $s \mathrel{|\!\approx} \alpha$
- $s \mathrel{|\!\approx} \mathbf{M}_{k+1}\, \alpha$ *iff* for some t and n, $s \cup \{t = n\} \mathrel{|\!\approx} \mathbf{M}_k\, \alpha$
- $s \mathrel{|\!\approx} \mathbf{O}\, \alpha$ *iff* s is minimal s.t. $s \mathrel{|\!\approx} \alpha$

Semantics in detail

- $s \approx (\neg)t = n$ *iff* $(\neg)t = n \in s$
- $s \approx (\alpha \vee \beta)$ *iff* $(\alpha \vee \beta) \in s$ or $s \approx \alpha$ or $s \approx \beta$
- $s \approx \neg(\alpha \vee \beta)$ *iff* $s \approx \neg\alpha$ and $s \approx \neg\beta$
- $s \approx \exists x\, \alpha$ *iff* $s \approx \alpha^x_n$ for some name n
- $s \approx \neg\exists x\, \alpha$ *iff* $s \approx \neg\alpha^x_n$ for every name n
- $s \approx \neg\neg\alpha$ *iff* $s \approx \alpha$

- $s \approx \mathbf{K}_0\, \alpha$ *iff* s is obviously inconsistent or $s \approx \alpha$
- $s \approx \mathbf{K}_{k+1}\, \alpha$ *iff* for some t and all n, $s \cup \{t = n\} \approx \mathbf{K}_k\, \alpha$
- $s \approx \mathbf{M}_0\, \alpha$ *iff* s is obviously consistent and $s \approx \alpha$
- $s \approx \mathbf{M}_{k+1}\, \alpha$ *iff* for some t and n, $s \cup \{t = n\} \approx \mathbf{M}_k\, \alpha$
- $s \approx \mathbf{O}\, \alpha$ *iff* s is minimal s.t. $s \approx \alpha$

obviously inconsistent $\hat{=}$ contains the empty clause
obviously consistent $\hat{=}$ not potentially inconsistent
potentially inconsistent $\hat{=}$

(a) obviously consistent
(b) two unsubsumed clauses mention two complementary literals
(c) for every name n, $t \neq n$ occurs in an unsubsumed clause

Semantics in detail

- $s_0, s, v \approx\!\!\!\!\models (\neg)t = n$ *iff* $(\neg)t = n \in s$
- $s_0, s, v \approx\!\!\!\!\models (\alpha \vee \beta)$ *iff* $(\alpha \vee \beta) \in s$ or $s_0, s, v \approx\!\!\!\!\models \alpha$ or $s_0, s, v \approx\!\!\!\!\models \beta$
- $s_0, s, v \approx\!\!\!\!\models \neg(\alpha \vee \beta)$ *iff* $s_0, s, v \approx\!\!\!\!\models \neg\alpha$ and $s_0, s, v \approx\!\!\!\!\models \neg\beta$
- $s_0, s, v \approx\!\!\!\!\models \exists x\, \alpha$ *iff* $s_0, s, v \approx\!\!\!\!\models \alpha^x_n$ for some name n
- $s_0, s, v \approx\!\!\!\!\models \neg\exists x\, \alpha$ *iff* $s_0, s, v \approx\!\!\!\!\models \neg\alpha^x_n$ for every name n
- $s_0, s, v \approx\!\!\!\!\models \neg\neg\alpha$ *iff* $s_0, s, v \approx\!\!\!\!\models \alpha$

- $s_0, s, v \approx\!\!\!\!\models \mathbf{K}_0\, \alpha$ *iff* $s_0 \cup v$ is obv. inconsistent or $s_0, s_0 \cup v, \emptyset \approx\!\!\!\!\models \alpha$
- $s_0, s, v \approx\!\!\!\!\models \mathbf{K}_{k+1}\, \alpha$ *iff* for some t and all n, $s_0, s, v \cup \{t = n\} \approx\!\!\!\!\models \mathbf{K}_k\, \alpha$
- $s_0, s, v \approx\!\!\!\!\models \mathbf{M}_0\, \alpha$ *iff* $s_0 \cup v$ is obv. consistent and $s_0, s_0 \cup v, \emptyset \approx\!\!\!\!\models \alpha$
- $s_0, s, v \approx\!\!\!\!\models \mathbf{M}_{k+1}\, \alpha$ *iff* for some t and n, $s_0, s, v \cup \{t = n\} \approx\!\!\!\!\models \mathbf{M}_k\, \alpha$
- $s_0, s, v \approx\!\!\!\!\models \mathbf{O}\, \alpha$ *iff* s_0 is minimal s.t. $s_0, s_0, \emptyset \approx\!\!\!\!\models \alpha$

obviously inconsistent $\hat{=}$ contains the empty clause
obviously consistent $\hat{=}$ not potentially inconsistent
potentially inconsistent $\hat{=}$

(a) obviously consistent
(b) two unsubsumed clauses mention two complementary literals
(c) for every name n, $t \neq n$ occurs in an unsubsumed clause