Hybrid-Logical Proofs: With an Application to the Sally-Anne Test

Torben Braüner
Roskilde University
Denmark

May 22, 2014

Plan of talk

I Brief introduction to hybrid logic

II Introduction to natural deduction

III Seligman's natural deduction system for hybrid logic

IV Application to the Sally-Anne test

V Concluding remarks

Part I

Brief introduction to hybrid logic

Arthur Prior (1914–1969)
The founding father of temporal logic
and what is now called hybrid logic

First hybrid-logical extension of ordinary modal logic:

Add a second sort of propositional symbols called *nominals*:

a, b, c, ...

Each nominal is true at exactly one time, thus, a nominal refers to a time.

First hybrid-logical extension of ordinary modal logic:

Add a second sort of propositional symbols called *nominals*:

a, b, c, . . .

Each nominal is true at exactly one time, thus, a nominal refers to a time.

A nominal can be used to formalize the example statement:

It is five o'clock 10 May 2007.

Or if times are replaced by persons:

I am Peter.

Like Arthur Prior's *egocentric logic*

Second hybrid-logical extension of ordinary modal logic:

Add operators called *satisfaction operators*:

$@_a$, $@_b$, $@_c$, ...

A formula $@_a\phi$ is true iff ϕ is true at the time a refers to.

Second hybrid-logical extension of ordinary modal logic:

Add operators called *satisfaction operators*:

$@_a$, $@_b$, $@_c$, ...

A formula $@_a\phi$ is true iff ϕ is true at the time a refers to.

The formula $@_a\phi$ can be used to formalize the statement:

At five o'clock 10 May 2007, it is raining.

Or if times are replaced by persons:

Peter is running.

Part II

Introduction to natural deduction

Natural deduction, cf. textbook by Warren Goldfarb

What we shall present is a system for deductions, *sometimes called a system of* natural deduction, *because to a certain extent it mimics certain natural ways we reason informally.*

*In particular, at any stage in a deduction we may introduce a new premise (that is, a new supposition); we may then infer things from this premise and eventually eliminate the premise (*discharge *it).*

Natural deduction rules for propositional logic

$$\frac{\phi \qquad \psi}{\phi \wedge \psi}\ (\wedge I) \qquad\qquad \frac{\phi \wedge \psi}{\phi}\ (\wedge E1) \qquad \frac{\phi \wedge \psi}{\psi}\ (\wedge E2)$$

$$\frac{\begin{array}{c}[\phi] \\ \vdots \\ \psi\end{array}}{\phi \to \psi}\ (\to I) \qquad\qquad \frac{\phi \to \psi \qquad \phi}{\psi}\ (\to E)$$

$$\frac{\begin{array}{c}[\neg\phi] \\ \vdots \\ \bot\end{array}}{\phi}\ (\bot 1)^{*}$$

$*$ ϕ is a propositional symbol ($\neg\phi$ is an abbreviation for $\phi \to \bot$)

Part III

Seligman's natural deduction system for hybrid logic

Rules for propositional logic and the following (modal operators are ignored)

Introduction and elimination rules for the satisfaction operator

$$\frac{a \qquad \phi}{@_a\phi}\ (@I) \qquad\qquad \frac{a \qquad @_a\phi}{\phi}\ (@E)$$

The (*Name*) rule gives a new name to the actual time (reflected in soundness proof)

$$\dfrac{\begin{array}{c}[a]\\ \vdots\\ \psi\end{array}}{\psi}\,(\textit{Name})^{\star}$$

$\star$ *a* does not occur free in ψ or in any undischarged assumptions other than the specified occurrences of *a*.

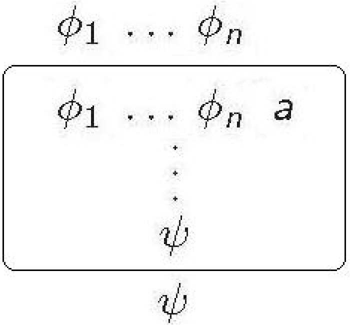
The (*Term*) rule enables hypothetical reasoning about what is the case at a specific time, possibly different from the actual time (reflected in soundness proof)

$$\cfrac{\phi_1 \quad \dots \quad \phi_n \qquad \begin{array}{c}[\phi_1]\dots[\phi_n][a]\\ \vdots \\ \psi\end{array}}{\psi}\;(\mathit{Term})^*$$

$*$ ϕ_1, ..., ϕ_n, and ψ are all satisfaction statements and there are no undischarged assumptions in the derivation of ψ besides the specified occurrences of ϕ_1, ..., ϕ_n, and a.

The way (*Term*) delimits a subderivation is similar to boxes in linear logic and the ($\Box I$) rule in some modal-logical proof-systems

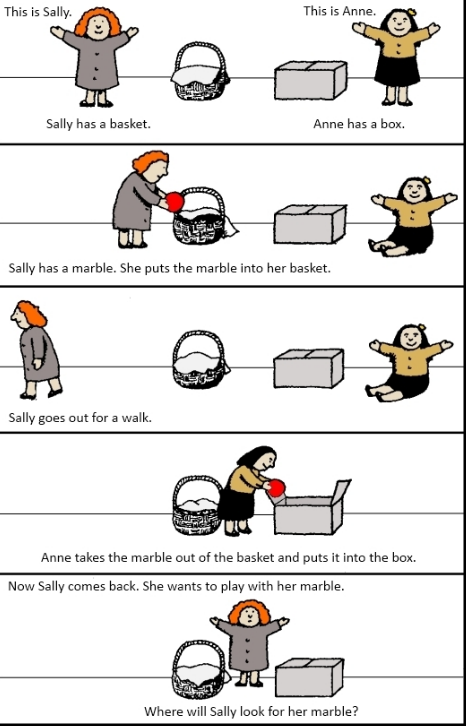
Alternative syntax of (*Term*) like boxes in linear logic



Cf also Blackburn, Braüner, Bolander and Jørgensen's work on Seligman-style tableus

Part IV

Application to the Sally-Anne test



The Sally-Anne test
measures a child's capacity
to ascribe false beliefs to others

Goes back to
Wimmer and Perner (1983)

Most children above the age of four
give the correct answer

Baron-Cohen, Leslie, and Frith (1985)
showed that autistic children have
a delayed ability to answer correctly

Answering the question correctly

We consider three successive times:
t_0 is before Sally leaves the scene
t_1 is when the marble is moved
t_2 is after Sally has returned

To answer the question, the child, say Peter, imagines himself being Sally, and he reasons as follows:

At t_0 Sally believes the marble is in the basket since she sees it, and she sees no action to move it, so at t_1, she also believe the marble is in the basket.

At t_2 she still believe that the marble is in the basket since she has not seen Anne moving it at the time t_1.

Therefore, Peter concludes that Sally at t_2 believes that the marble is in the basket.

We want to formalize the reasoning in the Sally-Anne test

Main assumption of our work:

Giving a correct answer to the Sally-Anne test involves a shift to the perspective of a different person, namely Sally, and back

We want to formalize this reasoning with local perspectives:

- Local perspectives can be represented by points in a Kripke model of modal logic
- Satisfaction operators can effect "jumps" between local perspectives

A principle of "inertia"

We need a principle of inertia saying that a belief is preserved over time, unless there is belief to the contrary

To this end we make use a standard successor-state axiom from Artificial Intelligence

$$(P2) \qquad Bp(t) \wedge \neg Bm(t) \rightarrow Bp(t+1)$$

with the following symbolizations

$p(t)$	The marble is in the basket at the time t
$m(t)$	The marble is moved from the basket at the time t

(Successor-state axioms are used to solve the frame problem)

Further principles

We also use the principles

(D)	$B\phi \rightarrow \neg B \neg \phi$
$(P1)$	$S\phi \rightarrow B\phi$
$(P3)$	$Bm(t) \rightarrow Sm(t)$

with the following symbolizations

S	Sees that ...
B	Believes that ...

Let the nominal a stand for the person Sally, then the correct answer can be formalized as

$$
\dfrac{@_a Sp(t_0)\ @_a S\neg m(t_0)\ @_a \neg Sm(t_1) \qquad \dfrac{[a] \qquad \dfrac{\dfrac{\dfrac{\dfrac{[a]\ [@_a Sp(t_0)]}{Sp(t_0)}}{Bp(t_0)}\,(P1) \qquad \dfrac{\dfrac{\dfrac{[a]\ [@_a S\neg m(t_0)]}{S\neg m(t_0)}}{B\neg m(t_0)}\,(P1)}{\neg Bm(t_0)}\,(\mathrm{D})}{Bp(t_1)}\,(P2) \qquad \dfrac{\dfrac{[a]\ [@_a \neg Sm(t_1)]}{\neg Sm(t_1)}}{\neg Bm(t_1)}\,(P3)}{Bp(t_2)}\,(P2)}{@_a Bp(t_2)}}{@_a Bp(t_2)}\,(Term)
$$

Note how the application of the rule ($Term$), marked in red, delimits the hypothetical reasoning taking place in a

Other work on logical analysis of the Sally-Anne test

Stenning and van Lambalgen (2008) analyse the Sally-Anne test (and other false-belief tasks) in terms of closed-world reasoning

Arkoudas and Bringsjord (2008) implement the reasoning in the Sally-Anne test in an interactive theorem prover

One major difference is the amount of first-order machinery:

	Our work	S. and van L.	A. and B.
Terms referring to times	Yes	Yes	Yes
Quantification over times	No	Yes	Yes
Quantification over events	No	No	Yes
Quantification over fluents	No	No	Yes

Summing up

Why is a *natural deduction* system for *hybrid modal logic* suitable for formalizing the Sally-Anne test?

- Natural deduction style proofs are meant to formalize the way human beings actually reason
- In modal logic, formulas are evaluated relative to points, representing local perspectives
- In hybrid logic it is possible to directly refer to such points, whereby local perspectives can be handled explicitly

Thus, hybrid-logical machinery can handle explicitly the different perspectives in the Sally-Anne test

Part V

Concluding remarks

Future work

- Different hypothetical reasoning
- Formalize other cognitive tasks
- Put a notion of identity on proofs to work
  1. When do two seemingly dissimilar reasoning tasks have the same underlying logical structure?
  2. When do two reasoning tasks have different logical structure, despite similarity?

  Might give rise to empirical predictions: If two dissimilar reasoning tasks have the same logical structure, then we would expect comparable empirical results
- More speculatively, contribute to the debate between the theory-theory and simulation-theory views of theory of mind

More information

The topic of this talk:

Braüner's paper Hybrid-Logical Reasoning in False-Belief Tasks in *Proceedings of Fourteenth conference on Theoretical Aspects of Rationality and Knowledge (TARK 2013)*

Hybrid logic in general:

Areces and ten Cate's chapter on hybrid logic in *Handbook of Modal Logic*, Elsevier, 2007

Braüner's chapter on hybrid logic in *Handbook of Philosophical Logic*, volume 17, Springer, 2013

Braüner's book *Hybrid Logic and its Proof-Theory*, Springer, 2011