Hybrid-Logical Proofs: With an Application to False-Belief Tasks

Torben Braüner
Roskilde University
Denmark

December 5, 2013

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 Smarties task

V What goes wrong when incorrect answers are given?

VI First-person vs third-person mental attribution

VII 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

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

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

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

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

Example: Natural deduction proof of $(p \wedge (p \to q)) \to q$

$$\dfrac{\dfrac{\dfrac{[p \wedge (p \to q)]}{p \to q}(\wedge E2) \qquad \dfrac{[p \wedge (p \to q)]}{p}(\wedge E1)}{q}(\to E)}{(p \wedge (p \to q)) \to q}(\to I)$$

Note how the formula $p \wedge (p \to q)$ is discharged

Part III

Seligman's natural deduction system for hybrid logic

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

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

The rules $(@I)$ and $(@E)$ formalizes the two informal arguments

It is Christmas Eve 2011;
it is snowing,
so at Christmas Eve 2011 it is snowing

It is Christmas Eve 2011;
at Christmas Eve 2011 it is snowing,
so it is snowing

$$\begin{array}{c} [a] \\ \vdots \\ \psi \\ \hline \psi \end{array} (\mathit{Name})^{\star}$$

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

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

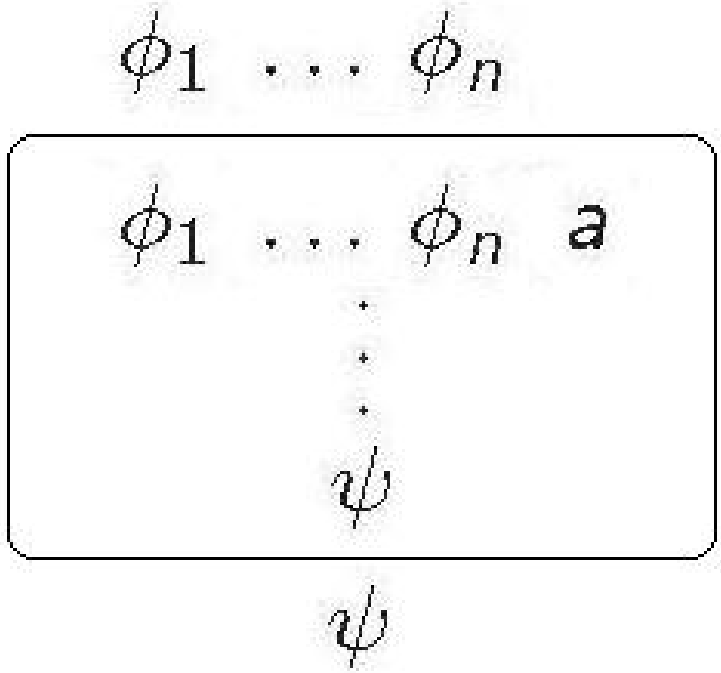
$$\frac{\phi_1 \quad \ldots \quad \phi_n \qquad\qquad \begin{array}{c}[\phi_1]\ldots[\phi_n][a]\\ \vdots \\ \psi\end{array}}{\psi}\,(\mathit{Term})^*$$

$*$ $\phi_1, \ldots, \phi_n$, and ψ are all satisfaction statements and there are no undischarged assumptions in the derivation of ψ besides the specified occurrences of $\phi_1, \ldots, \phi_n$, and 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)

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



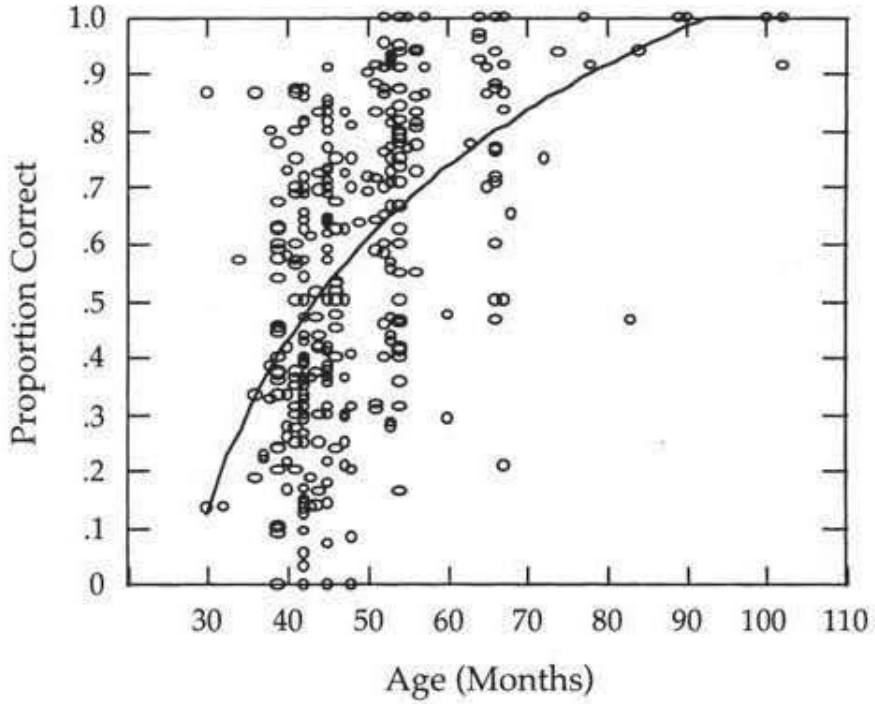
Cf also Blackburn, Braüner, Bolander and Jørgensen (2013) on Seligman-style tableus

Part IV

Application to the Smarties task

Several different false-belief tasks

Shows a very robust pattern:



Demonstrated in the meta-analysis Wellman et al. (2001) involving 178 individual false-belief studies and more than 4000 children

Autistic children have a delayed ability to answer correctly

The Smarties task

A child, say Peter, is shown a Smarties tube where unbeknownst to him the Smarties have been replaced by pencils.

Peter is asked (first question):

"What do you think is inside the tube?"

Peter answers:

"Smarties!"

The tube is then shown to contain pencils only.

Peter is then asked (second question):

"Before it was opened, what did you think was inside?"

Peter answers:

"Before it was opened, I thought there were ??? inside"

(Comes in a person version as well, used to test Theory of Mind)

Answering the second question correctly

Peter imagines himself being at the earlier time, say *a*, where he was asked the first question.

At that time he deduced that there were Smarties inside from the fact that it is a Smarties tube.

Imagining being at the time *a*, Peter reasons that since he at that time deduced that there were Smarties inside, he must also have come to believe this.

Therefore, he concludes that at the earlier time *a* he believed that there were Smarties inside.

We want to formalize the reasoning in the Smarties task

Main assumption of our work:

Giving a correct answer to the Smarties task involves a shift to a different perspective (time or person), 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

The correct answer to the second question, more formally

Premise 1:

At the time a, Peter deduces that there are Smarties inside the tube

Premise 2:

If Peter deduces ϕ then Peter believes ϕ

Conclusion:

At the time a, Peter believes that there are Smarties inside the tube

The deduction step from the premises to the conclusion involves a shift of perspective to the hypothetical time *a*, and back

(This hypothetical reasoning is necessary to reach the conclusion)

We use the following symbolization

s	There are Smarties inside the tube
a	The time where the first question is asked
D	Peter deduces that ...
B	Peter believes that ...

and take the principle $D\phi \to B\phi$ as an axiom schema

Then the correct answer can be completely formalized as

$$\dfrac{@_a Ds \qquad \dfrac{\dfrac{[a] \qquad [@_a Ds]}{Ds}(@E) \qquad \dfrac{}{Ds \to Bs}}{Bs}(\to E)}{@_a Bs}(@I)}{@_a Bs}(Term)$$

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

Summing up

Why is a *natural deduction* system for *hybrid modal logic* suitable for formalizing the Smarties task?

- 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 Smarties task

Part V

What goes wrong when incorrect answers are given?

A pattern in the incorrect answers (normative vs descriptive)

The derivation of the correct answer given earlier does not explicitly tell what goes wrong when an incorrect answer is given.

But a child either answers correctly, or gives a specific incorrect answer.

In case of the Smarties task, the child (incorrectly) answers "Pencils", not something else irrelevant.

We will look for a pattern in the incorrect answers.

Now, consider two stages in the correct answer

Peter is shown the Smarties tube where the Smarties have been replaced by pencils. Peter is asked question one:

> *"What do you think is inside the tube?"*

Peter answers:

> *"Smarties!"*

The tube is then shown to contain pencils only.

Stage 1: The actual content has been disclosed and Peter has come to believe that there are pencils inside ($@_{\mathtt{peter}} Bp$)

Peter is then asked the second question:

> *"If your mother comes into the room and we show this tube to her, what will she think is inside?"*

Stage 2: Question two has been asked and Peter derives that the mother believes that there are Smarties inside ($@_{\mathtt{mother}} Bs$)

Peter answers:

> *My mother will think that there are Smarties inside*

Formalization of information at Stage 1

We have used the following additional symbolization

p	There are pencils inside the tube
$\mathtt{peter}$	The person Peter
S	Sees that ...

At Stage 1, Peter sees that the tube contains pencils ($@_{\mathtt{peter}} Sp$), so he comes to believe that there are pencils inside ($@_{\mathtt{peter}} Bp$)

Taking $S\phi \rightarrow B\phi$ as an axiom schema, this can be formalized as

$$\dfrac{\dfrac{\mathtt{peter} \qquad @_{\mathtt{peter}} Sp}{Sp}\,(@E) \qquad \dfrac{}{Sp \rightarrow Bp}\,(\rightarrow E)}{\dfrac{\mathtt{peter} \qquad Bp}{@_{\mathtt{peter}} Bp}\,(@I)}$$

Note that the nominal $\mathtt{peter}$ is true as it is Peter that performs the reasoning

Note also that no perspective shift is needed!

Formalization of information at Stage 2 (correct reasoning)

At Stage 2, Peter tries to figure out what the mother believes is inside the tube ($@_{\texttt{mother}} B???$)

Analogous to the temporal case described earlier, Peter reasons that the mother must deduce that the tube contains Smarties ($@_{\texttt{mother}} Ds$), and from that, she must also come to believe that there are Smarties inside ($@_{\texttt{mother}} Bs$).

Peter therefore answers:

> *My mother will think that there are Smarties inside* ($@_{\texttt{mother}} Bs$)

(Leaving the information obtained at Stage 1 irrelevant)

Summary of correct reasoning

Information obtained at Stage 1:
Peter believes that there are pencils inside ($@_{\mathtt{peter}} Bp$)

Question: $@_{\mathtt{mother}} B???$

Information obtained at Stage 2:
The mother believes that there are Smarties inside ($@_{\mathtt{mother}} Bs$)

Correct answer: $@_{\mathtt{mother}} Bs$

What goes wrong when an incorrect answer is given?

Information obtained at Stage 1:
Peter believes that there are pencils inside ($@_{\mathtt{peter}} Bp$)

Question: $@_{\mathtt{mother}} B???$

Information obtained at Stage 2: None

Incorrect answer: $@_{\mathtt{mother}} Bp$

Thus, instead of giving the correct answer, reporting what the mother believes is inside the tube, Peter reports what he himself believes is inside

(Making use of the information obtained at Stage 1)

A pattern of failure

The child giving an incorrect answer does not perform the shift of perspective required to be able to figure out what the mother believe is the case

Instead the child reports what is believed to be the case from the childs own perspective

Thus, this "pattern of failure" gives a formal corroboration of the claim that children under four and autistic children have difficulties shifting to a perspective different from their own.

(Same pattern in case of the Sally-Anne task)

Part VI

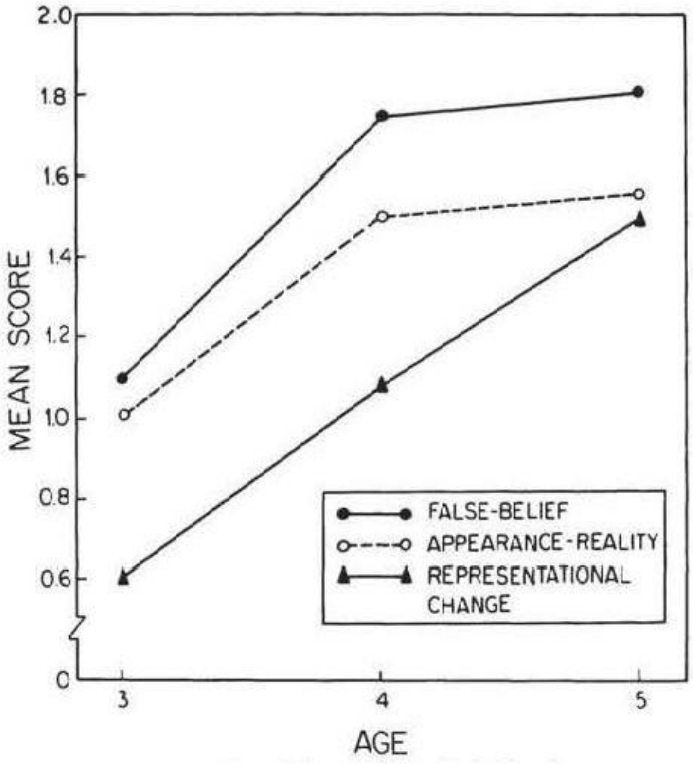
First-person vs third-person mental attribution

We have considered two versions of the Smarties task

- Peter shifts perspective to an earlier time where a false belief is attributed to himself (first-person)
- Peter shifts perspective to the mother who is attributed a false belief (third-person)

One should think that being asked about someone else's belief is different from being asked about one's own

But third- and first-person mental attribution develop in parallel



Observed in Gopnik and Astington (1988)
and confirmed by the meta-analysis Wellman et al. (2001)

Is there a common cognitive basis? Extensively debated...

Same logical structure

Now, we have demonstrated formally that the two versions of the Smarties task have the same logical structure

Is this an explanation of the parallel empirical results?

(Rather than a common cognitive basis?)

Part VII

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