Proof theory for deontic logic inspired by Indian Philosophy

Björn Lellmann

TU Wien

Annual meeting of the SSLPS
Bern, Oct 25, 2019

What's the problem?

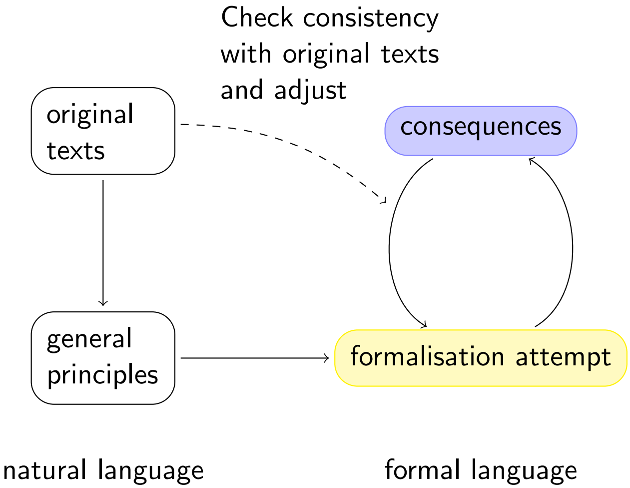
Main idea:

Use logic to formally analyse ancient texts of Indian Philosophy

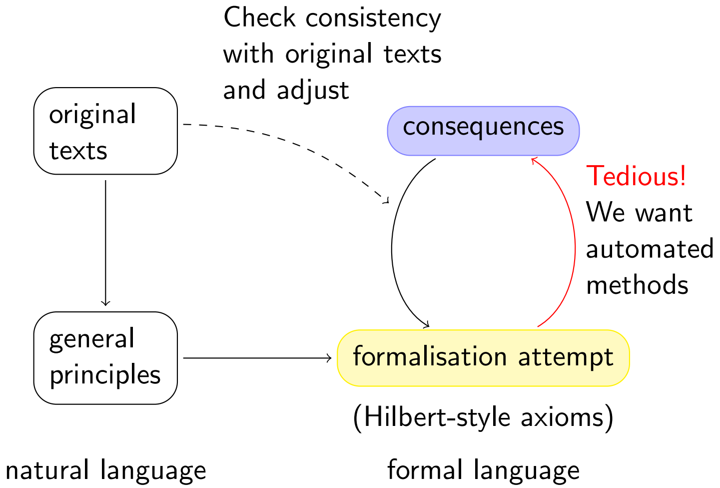
Expected benefits:

- Indology: A better understanding of the texts and clarification through formalisation
- Logic: New inputs and development of new methods

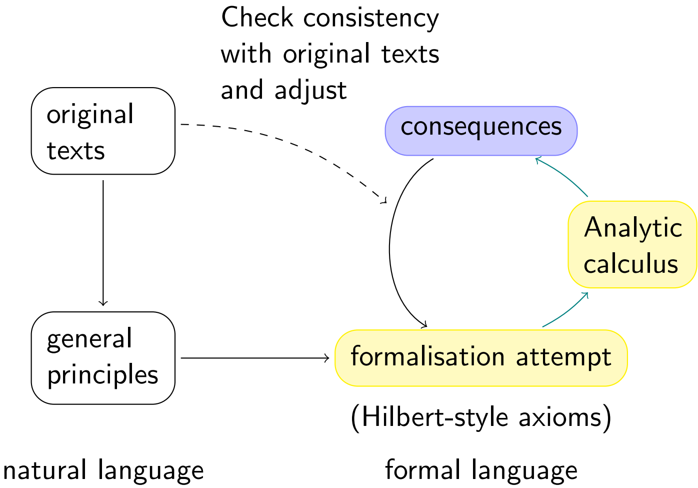
How do we formalise?



How do we formalise?



How do we formalise?



So we use Proof Theory to do the dirty work!

What kind of Indian Philosophy?

We consider texts of the Mīmāṃsā school:

- Main period of activity
  last centuries BCE to beginning of 20th century

- Main focus
  interpretation of the prescriptive portions of the Vedas

- Main tool
  formulation of general rules and interpretative principles (nyāyas) to interpret the texts

What kind of Indian Philosophy?

We consider texts of the Mīmāṃsā school:

- Main period of activity
  last centuries BCE to beginning of 20th century

- Main focus
  interpretation of the prescriptive portions of the Vedas
  $\rightsquigarrow$ Deontic Logic
- Main tool
  formulation of general rules and interpretative principles (nyāyas) to interpret the texts
  $\rightsquigarrow$ Axioms

But why this kind of Indian Philosophy – Why Mīmāṃsā?

- Suitability: The clear formulation of Mīmāṃsā interpretative principles lends itself to formalisation
- Historical significance: Mīmāṃsā is one of the main schools of Indian Philosophy and considered early deontic logic
- Importance: The Mīmāṃsā principles are used in Indian court cases even today
- Novelty: Mīmāṃsā texts have scarcely been considered from the modern Western point of view due to:
    - Lack of translations
    - Often highly metaphorical language

Part 1: The logic – formalising the key concepts

Preliminaries: The language

How to model propositional reasoning?

> *When there is a contradiction, at the denial of one alternative, the other is known (to be true).*

(Interpretation of Jayanta's Nyāyamañjarī, 9^{th} c. CE)

$\rightsquigarrow$ classical propositional logic

How to model the deontic concepts?

Mīmāṃsā authors use eligibility conditions for prescriptions, e.g.

> *The one who desires heaven should sacrifice with the Full- and New-moon Sacrifices*

$\rightsquigarrow$ dyadic deontic operators

Note: We base our logic on propositions instead of actions.

Preliminaries: The language, formally

The formulae of deontic logic are given by:

$$p \in \mathsf{Var} \mid \bot \mid \neg A \mid A \wedge B \mid A \vee B \mid A \to B \mid \mathcal{O}(A/B) \mid \mathcal{F}(A/B)$$

They are interpreted in the standard way:

- $\mathcal{O}(A/B) \rightsquigarrow$ "Given that B is the case, it is obligatory that A is the case"
- $\mathcal{F}(A/B) \rightsquigarrow$ "Given that B is the case, it is forbidden that A is the case"

As base calculus we assume the (Hilbert-)axioms and rules of classical propositional logic and the congruence rules

$$\frac{\vdash A \leftrightarrow C \quad \vdash B \leftrightarrow D}{\vdash \mathcal{O}(A/B) \leftrightarrow \mathcal{O}(C/D)} \qquad \frac{\vdash A \leftrightarrow C \quad \vdash B \leftrightarrow D}{\vdash \mathcal{F}(A/B) \leftrightarrow \mathcal{F}(C/D)}$$

The logic: Axioms

> *When, on the other hand, coming into being [of something needed] [...] are not realized by another prescription, [the principal prescription] itself begets the four [stages] of coming into being [...] [of the prescriptions] connected to itself.*

(Rāmānujācārya, *Tantrahasya* IV.4.3.3)

The logic: Axioms

When, on the other hand, coming into being [of something needed] [...] are not realized by another prescription, [the principal prescription] itself begets the four [stages] of coming into being [...] [of the prescriptions] connected to itself.

(Rāmānujācārya, *Tantrahasya* IV.4.3.3)

If a prescription enjoins something which has requirements, then it enjoins the requirements as well.

$$\frac{\vdash A \rightarrow B}{\vdash \mathcal{O}(A/C) \rightarrow \mathcal{O}(B/C)}$$

The logic: More axioms

Given that purposes Y and Z exclude each other, if one should use X for the purpose Y, then it cannot be the case that one should use it at the same time for the purpose Z.

(Interpretation of Kumārila, *Tantravārttika* on PMS 1.3.3)

$$\rightsquigarrow \qquad \frac{\vdash \neg(A \wedge B)}{\vdash \neg(\mathcal{O}(A/C) \wedge \mathcal{O}(B/C))}$$

If conditions X and Y are always equivalent, given the duty to perform Z under conditions X, the same duty applies under Y.

(Interpretation of Śabara, PMS 6.1.25)

$$\rightsquigarrow \qquad \frac{\vdash B \leftrightarrow C}{\vdash \mathcal{O}(A/B) \rightarrow \mathcal{O}(A/C)}$$

Making the logic useful: Sequents

These rules are equivalent to dyadic deontic logic MD:

$(M)\ \ \mathcal{O}(A \wedge B/C) \rightarrow \mathcal{O}(A/C)$

$(D)\ \ \neg(\mathcal{O}(A/B) \wedge \mathcal{O}(\neg A/B))$

$$\frac{\vdash A \leftrightarrow C \quad \vdash B \leftrightarrow D}{\vdash \mathcal{O}(A/B) \leftrightarrow \mathcal{O}(C/D)}\ Cg$$

Making the logic useful: Sequents

These rules are equivalent to dyadic deontic logic MD:

(M) $\mathcal{O}(A \wedge B/C) \rightarrow \mathcal{O}(A/C)$

(D) $\neg(\mathcal{O}(A/B) \wedge \mathcal{O}(\neg A/B))$

$$\frac{\vdash A \leftrightarrow C \quad \vdash B \leftrightarrow D}{\vdash \mathcal{O}(A/B) \leftrightarrow \mathcal{O}(C/D)}\ Cg$$

(Turning the handle to
get analytic calculus)

Making the logic useful: Sequents

These rules are equivalent to dyadic deontic logic MD:

$(M)\ \ \mathcal{O}(A \wedge B/C) \to \mathcal{O}(A/C)$

$(D)\ \ \neg(\mathcal{O}(A/B) \wedge \mathcal{O}(\neg A/B))$

$$\frac{\vdash A \leftrightarrow C \qquad \vdash B \leftrightarrow D}{\vdash \mathcal{O}(A/B) \leftrightarrow \mathcal{O}(C/D)}\ Cg$$

(Turning the handle to get analytic calculus)

$$\frac{A \Rightarrow C \qquad B \Rightarrow D \qquad D \Rightarrow B}{\Gamma, \mathcal{O}(A/B) \Rightarrow \mathcal{O}(C/D), \Delta}\ \mathsf{Mon}_1$$

$$\frac{A, C \Rightarrow \qquad B \Rightarrow D \qquad D \Rightarrow B}{\Gamma, \mathcal{O}(A/B), \mathcal{O}(C/D) \Rightarrow \Delta}\ \mathsf{D}$$

$$\frac{A \Rightarrow}{\Gamma, \mathcal{O}(A/B) \Rightarrow \Delta}\ \mathsf{P}$$

A sequent calculus suitable for automatic proof search!

(A sequent $A_1, \ldots, A_n \Rightarrow B_1, \ldots, B_m$ reads as $\bigwedge_{i=1}^{n} A_i \to \bigvee_{j=1}^{m} B_j$)

...so let's try to use the calculus!

Suppose that

- Śūdras must not study the Vedas: $\mathcal{O}(\neg\texttt{std\_vds}/\texttt{sdr})$
- Performing the Agnihotra sacrifice demands studying the Vedas: $\texttt{agnhtr} \to \texttt{std\_vds}$

Question 1: May Śūdras perform Agnihotra?

$$\dfrac{\dfrac{\texttt{agnhtr} \overset{\checkmark}{\Rightarrow} \texttt{std\_vds}}{\neg\texttt{std\_vds} \Rightarrow \neg\texttt{agnhtr}} \quad \dfrac{\checkmark}{\texttt{sdr} \Rightarrow \texttt{sdr}} \quad \dfrac{\checkmark}{\texttt{sdr} \Rightarrow \texttt{sdr}}}{\mathcal{O}(\neg\texttt{std\_vds}/\texttt{sdr}) \Rightarrow \mathcal{O}(\neg\texttt{agnhtr}/\texttt{sdr})}\ \text{Mon}_1$$

Answer: No, they must not!

. . . so let's try to use the calculus!

Suppose that

- Śūdras must not study the Vedas: $\mathcal{O}(\neg\texttt{std\_vds}/\texttt{sdr})$
- Performing the Agnihotra sacrifice demands studying the Vedas: $\texttt{agnhtr} \rightarrow \texttt{std\_vds}$
- Chariot makers are Śūdras: $\texttt{chmk} \rightarrow \texttt{sdr}$

Question 2: What about chariot makers?

$$\dfrac{\dfrac{\checkmark}{\neg\texttt{std\_vds} \Rightarrow \neg\texttt{std\_vds}} \quad \overset{X}{\texttt{sdr} \Rightarrow \texttt{chmk}} \quad \overset{\checkmark}{\texttt{chmk} \Rightarrow \texttt{sdr}}}{\mathcal{O}(\neg\texttt{std\_vds}/\texttt{sdr}) \Rightarrow \mathcal{O}(\neg\texttt{std\_vds}/\texttt{chmk})}\ \mathsf{Mon}_1$$

Answer: We can't derive anything!
(Because the logic is too weak in the second argument.)

Part 2: Reasoning on conditions

How to reason on the conditions?

We cannot introduce full downwards monotony for conditions: $\mathcal{O}(\neg\texttt{std\_vds}/\texttt{sdr})$ and $\mathcal{O}(\texttt{agnhtr}/\texttt{chmk})$ would give

$$\mathcal{O}(\neg\texttt{std\_vds}/\texttt{chmk}) \wedge \mathcal{O}(\texttt{std\_vds}/\texttt{chmk})$$

which is inconsistent with axiom (D).

So we distinguish (prima-facie) assumptions from derived statements and use:

> *Guṇapradhāna / specificity principle*
> *More specific rules override more general ones.*

Discussed by Jaimini (2^{nd} c. BCE)

To make this precise we split the assumptions into:

- propositional assumptions (facts): $\bigwedge_{i \leq n} p_i \rightarrow \bigvee_{j \leq m} p_j$
- deontic assumptions: $\mathcal{O}_{\text{pf}}(A/B)$ with A, B propositional.

Guṇapradhāna / Specificity intuitively

Idea: Use downwards monotonicity in the second argument...

$\texttt{agnhtr} \rightarrow \texttt{std\_vds}$ $\quad$ $\texttt{chmk} \rightarrow \texttt{sdr}$

$\mathcal{O}_{\mathsf{pf}}(\neg\texttt{std\_vds}/\texttt{sdr})$

$\mid\!\sim \mathcal{O}(\neg\texttt{std\_vds}/\texttt{chmk} \wedge \neg\texttt{upnyn})$

Guṇapradhāna / Specificity intuitively

Idea: Use downwards monotonicity in the second argument if the obligation is not overruled by a more specific one…

$\texttt{agnhtr} \rightarrow \texttt{std\_vds}$ $\qquad$ $\texttt{chmk} \rightarrow \texttt{sdr}$

$\mathcal{O}_{\mathsf{pf}}(\neg\texttt{std\_vds}/\texttt{sdr})$

$\lightning$

$\mathcal{O}_{\mathsf{pf}}(\texttt{agnhtr}/\texttt{chmk})$

$\not\mid\!\sim\ \mathcal{O}(\neg\texttt{std\_vds}/\texttt{chmk} \wedge \neg\texttt{upnyn})$

$\mid\!\sim\ \mathcal{O}(\ \texttt{std\_vds}/\texttt{chmk} \wedge \neg\texttt{upnyn})$

Guṇapradhāna / Specificity intuitively

Idea: Use downwards monotonicity in the second argument
if the obligation is not overruled by a more specific one
and if there is no other conflicting one which is not overruled itself.

$$\mathtt{agnhtr} \rightarrow \mathtt{std\_vds} \qquad \mathtt{chmk} \rightarrow \mathtt{sdr}$$

$$\mathcal{O}_{\mathsf{pf}}(\neg \mathtt{std\_vds} / \mathtt{sdr})$$

$$\lightning$$

$$\mathcal{O}_{\mathsf{pf}}(\neg \mathtt{std\_vds} / \neg \mathtt{upnyn})$$

$$\mathcal{O}_{\mathsf{pf}}(\mathtt{agnhtr} / \mathtt{chmk})$$

$$\not\models \mathcal{O}(\neg \mathtt{std\_vds} / \mathtt{chmk} \wedge \neg \mathtt{upnyn})$$

$$\not\models \mathcal{O}(\ \mathtt{std\_vds} / \mathtt{chmk} \wedge \neg \mathtt{upnyn})$$

How to put this into sequents?

We derive an obligation $\mathcal{O}(A/B)$ from deontic assumptions $\mathfrak{L}$ if

- it is entailed by an applicable $\mathcal{O}_{\mathsf{pf}}(C/D) \in \mathfrak{L}$...

$$\{B \Rightarrow D\} \quad \cup \quad \{C \Rightarrow A\}$$

$$\frac{}{\Rightarrow \mathcal{O}(A/B)}\ \mathcal{O}_R^{\mathcal{O}_{\mathsf{pf}}}$$

How to put this into sequents?

We derive an obligation $\mathcal{O}(A/B)$ from deontic assumptions $\mathfrak{L}$ if

- it is entailed by an applicable $\mathcal{O}_{\mathsf{pf}}(C/D) \in \mathfrak{L}$
  which is not overruled by a conflicting more specific one

$$\dfrac{\{B \Rightarrow D\} \quad \cup \quad \{C \Rightarrow A\} \qquad \left\{ \bigvee \begin{pmatrix} \{\nvdash B \Rightarrow F\} \\ \{\nvdash F \Rightarrow D\} \\ \{\nvdash E, A \Rightarrow \ \} \end{pmatrix} \mid \mathcal{O}_{\mathsf{pf}}(E/F) \in \mathfrak{L} \right\}}{\Rightarrow \mathcal{O}(A/B)} \; \mathcal{O}_R^{\mathcal{O}_{\mathsf{pf}}}$$

How to put this into sequents?

We derive an obligation $\mathcal{O}(A/B)$ from deontic assumptions $\mathfrak{L}$ if

- it is entailed by an applicable $\mathcal{O}_{\mathsf{pf}}(C/D) \in \mathfrak{L}$ which is not overruled by a conflicting more specific one
- there is no other applicable conflicting obligation

  ...

$$
\frac{\{B \Rightarrow D\} \quad \cup \quad \{C \Rightarrow A\} \qquad \left\{ \bigvee \begin{pmatrix} \{\nvdash B \Rightarrow F\} \\ \{\nvdash F \Rightarrow D\} \\ \{\nvdash E, A \Rightarrow \ \} \end{pmatrix} \mid \mathcal{O}_{\mathsf{pf}}(E/F) \in \mathfrak{L} \right\} \qquad \left\{ \bigvee \begin{pmatrix} \{\nvdash B \Rightarrow H\} \\ \{\nvdash G, A \Rightarrow \ \} \\ \\ \\ \end{pmatrix} \mid \mathcal{O}_{\mathsf{pf}}(G/H) \in \mathfrak{L} \right\}}{\Rightarrow \mathcal{O}(A/B)} \ \mathcal{O}_R^{\mathcal{O}_{\mathsf{pf}}}
$$

How to put this into sequents?

We derive an obligation $\mathcal{O}(A/B)$ from deontic assumptions $\mathfrak{L}$ if

- it is entailed by an applicable $\mathcal{O}_{\mathsf{pf}}(C/D) \in \mathfrak{L}$ which is not overruled by a conflicting more specific one
- there is no other applicable conflicting obligation which is not overruled itself by a more specific one

$$\dfrac{\{B \Rightarrow D\} \quad \cup \quad \{C \Rightarrow A\} \qquad \left\{ \bigvee \begin{pmatrix} \{\nvdash B \Rightarrow F\} \\ \{\nvdash F \Rightarrow D\} \\ \{\nvdash E, A \Rightarrow \ \} \end{pmatrix} \mid \mathcal{O}_{\mathsf{pf}}(E/F) \in \mathfrak{L} \right\} \qquad \left\{ \bigvee \begin{pmatrix} \{\nvdash B \Rightarrow H\} \\ \{\nvdash G, A \Rightarrow \ \} \\ \left\{ \begin{array}{l} \{B \Rightarrow J\} \\ \cup \{J \Rightarrow H\} \\ \cup \{I \Rightarrow A\} \end{array} \mid \mathcal{O}_{\mathsf{pf}}(I/J) \in \mathfrak{L} \right\} \end{pmatrix} \mid \mathcal{O}_{\mathsf{pf}}(G/H) \in \mathfrak{L} \right\}}{\Rightarrow \mathcal{O}(A/B)} \; \mathcal{O}_R^{\mathcal{O}_{\mathsf{pf}}}$$

Wait … underivability premisses?

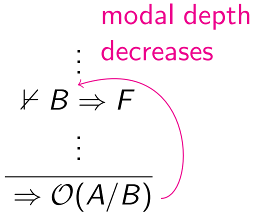
The underivability premisses look fishy and smell like circular definitions…

Fortunately:

Theorem.
Derivability for formulae of modal depth $n+1$ depends only on derivability of formulae of modal depth at most n.

So everything is well-defined, we escape a fixpoint definition and even get

Theorem.
Derivability from assumptions is decidable in polynomial space.



(with $\mathcal{O}_{\mathsf{pf}}(E/F) \in \mathfrak{L}$, hence no $\mathcal{O}_{\mathsf{pf}}$ in F)

Bonus: Saving the twins (well . . . at least one)

Vikalpa / Disjunctive response:
When there is a real conflict between obligations, any of the conflicting injunctions may be adopted as option.

Discussed by Jaimini (2^{nd} c. BCE)

This principle actually follows from the methods and lets us reason about the drowning twins:

$$\texttt{save\_twin1} \wedge \texttt{save\_twin2} \rightarrow \bot$$

$$\mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin1}/\texttt{reach1}) \qquad \mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin2}/\texttt{reach2})$$

$$\mid\!\sim \mathcal{O}(\texttt{save\_twin1} \;/\; \texttt{reach1})$$

Bonus: Saving the twins (well . . . at least one)

Vikalpa / Disjunctive response:

When there is a real conflict between obligations, any of the conflicting injunctions may be adopted as option.

Discussed by Jaimini (2^{nd} c. BCE)

This principle actually follows from the methods and lets us reason about the drowning twins:

$$\texttt{save\_twin1} \wedge \texttt{save\_twin2} \rightarrow \bot$$

↯

$$\mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin1}/\texttt{reach1}) \qquad \mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin2}/\texttt{reach2})$$

$$\not\vdash \mathcal{O}(\texttt{save\_twin2}\,/\,\texttt{reach1})$$

Bonus: Saving the twins (well . . . at least one)

Vikalpa / Disjunctive response:
When there is a real conflict between obligations, any of the conflicting injunctions may be adopted as option.

Discussed by Jaimini (2nd c. BCE)

This principle actually follows from the methods and lets us reason about the drowning twins:

$$\texttt{save\_twin1} \wedge \texttt{save\_twin2} \rightarrow \bot$$

$$\mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin1}/\texttt{reach1}) \qquad \mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin2}/\texttt{reach2})$$

$$\not\approx\ \mathcal{O}(\texttt{save\_twin1} \,/\, \texttt{reach1} \wedge \texttt{reach2})$$

Bonus: Saving the twins (well ... at least one)

Vikalpa / Disjunctive response:
When there is a real conflict between obligations, any of the conflicting injunctions may be adopted as option.

Discussed by Jaimini (2nd c. BCE)

This principle actually follows from the methods and lets us reason about the drowning twins:

$$\texttt{save\_twin1} \wedge \texttt{save\_twin2} \rightarrow \bot$$

$$\mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin1}/\texttt{reach1}) \qquad \mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin2}/\texttt{reach2})$$

$$\not\approx \mathcal{O}(\texttt{save\_twin2} \,/\, \texttt{reach1} \wedge \texttt{reach2})$$

Bonus: Saving the twins (well . . . at least one)

Vikalpa / Disjunctive response:

When there is a real conflict between obligations, any of the conflicting injunctions may be adopted as option.

Discussed by Jaimini (2nd c. BCE)

This principle actually follows from the methods and lets us reason about the drowning twins:

$$\texttt{save\_twin1} \wedge \texttt{save\_twin2} \rightarrow \bot$$

$$\mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin1}/\texttt{reach1}) \qquad \mathcal{O}_{\mathsf{pf}}(\texttt{save\_twin2}/\texttt{reach2})$$

$$\mid\!\sim\ \mathcal{O}(\texttt{save\_twin1} \vee \texttt{save\_twin2} \,/\, \texttt{reach1} \wedge \texttt{reach2})$$

Application: Comparing formalisations

The procedure also helps in comparing and adjusting formalisations of deontic assumptions using:

Theorem.
If $\mathcal{O}_{\mathsf{pf}}(A/B) \in \mathfrak{L}$, and $\mathcal{O}(A/B)$ is not derivable, then there is an explicit conflict: a $\mathcal{O}_{\mathsf{pf}}(C/D) \in \mathfrak{L}$ with $\vdash \neg(A \wedge C)$ and $\vdash B \leftrightarrow D$.

Explicit conflicts are bad:

A deontic assumption involved in an explicit conflict
is never applied as written!

This suggests for a list of deontic assumptions:

- count explicit conflicts in a conflict score
- if possible, generate alternative formalisations
- evaluate the formalisations by minimising the conflict score.

Application: Comparing interpretations

Implementation

Interpretation and formalisation

- $\mathcal{O}_{\text{pf}}(A/B)$
- $\mathcal{F}_{\text{pf}}(A/B)$

$\vdots$

$\mathcal{O}(A \vee \neg A/B)$

$\mathcal{F}(A \wedge \neg A/B)$

- Do A in B
- Don't do A in B

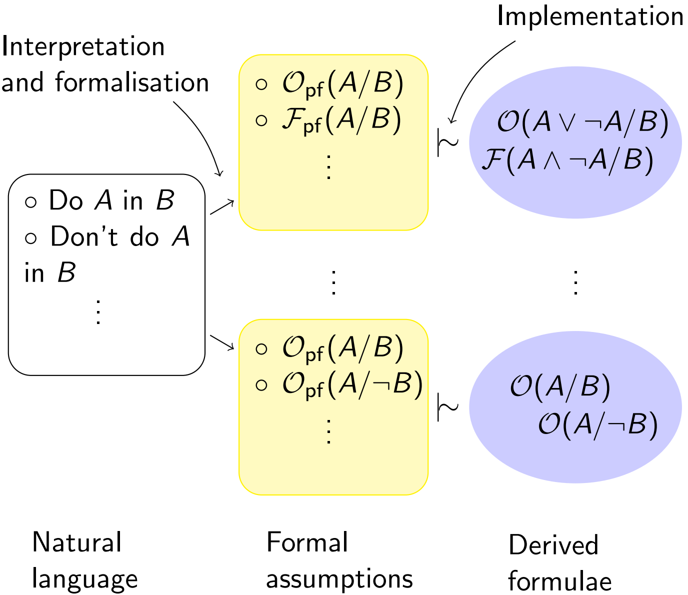
$\vdots$

Natural language

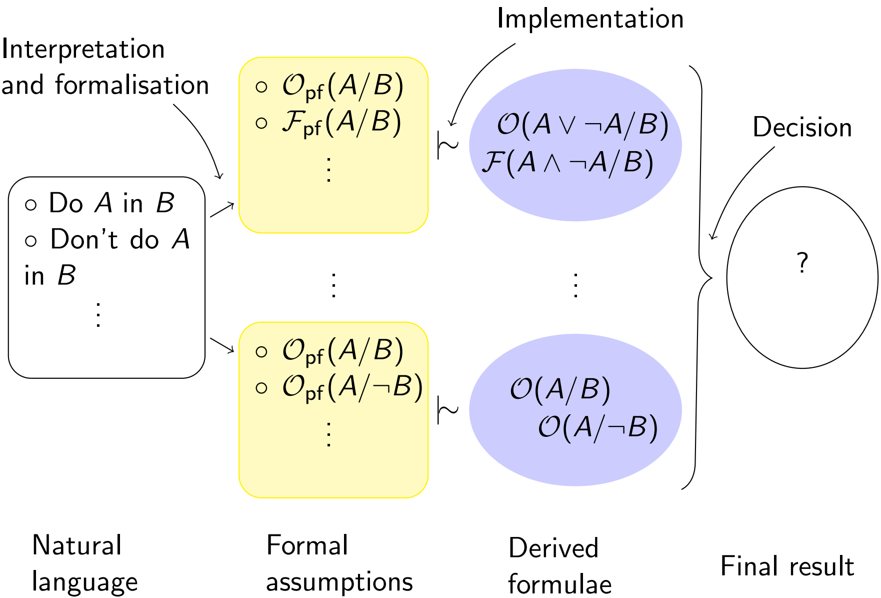
Formal assumptions

Derived formulae

Application: Comparing interpretations



Application: Comparing interpretations



What does this look like in practice?

Let the implementation work that out!

http://subsell.logic.at/bprover/deonticProver/version1.2/

Summing up

In this line of work we have

- investigated proof-theoretic methods for converting axiom systems into analytic (sequent) calculi
- Obtained a known deontic logic from texts of the Mīmāṃsā school of Indian Philosophy
- Introduced a proof-theoretic mechanism for reasoning with conflicting obligations using specificity
- Implemented a tool which also helps to evaluate the formalisation of deontic assumptions

Thank you!

`https://mimamsa.logic.at`