Permission in Deontic Logic: From Sanskrit Philosophy to AI

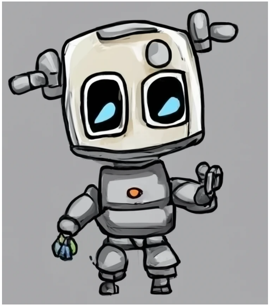
Josephine Dik

Supervisor: Agata Ciabattoni

Institute of Logic and Computation
Theory and Logic group

josephine@logic.at

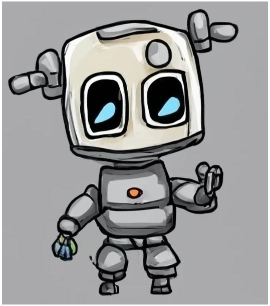
Autonomous Agent



Actions are constrained by **norms**

- Obligations?
- Prohibitions?
- Permissions?

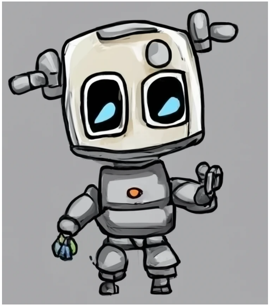
Autonomous Agent



Actions are constrained by **norms**

- Obligations! - Actions you have to do
- Prohibitions?
- Permissions?

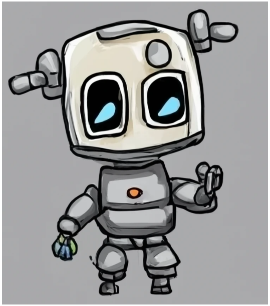
Autonomous Agent



Actions are constrained by **norms**

- Obligations! - Actions you have to do
- Prohibitions! - Actions you are not allowed to do
- Permissions?

Autonomous Agent



Actions are constrained by **norms**

- Obligations! - Actions you have to do
- Prohibitions! - Actions you are not allowed to do
- Permissions? - Unclear

Outline

1. Permission in Deontic Logic
2. Mīmāṃsā Deontic
3. Mīmāṃsā Permission
4. Incorperating Preferences

Outline

1 Permission in Deontic Logic

2 Mīmāṃsā Deontic

3 Mīmāṃsā Permission

4 Incorperating Preferences

Permission in Standard Deontic Logic (SDL)

Deontic Logic: reasoning about norms (obligations, prohibitions, permissions)

- Introduced by Von Wright, 1951:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid (\varphi \rightarrow \varphi) \mid \mathcal{P}\,\varphi$$

- φ is **permitted**: $\mathcal{P}\,\varphi$
- φ is **forbidden**: $\mathcal{F}\,\varphi := \neg\,\mathcal{P}\,\varphi$
- φ is **obligatory**: $\mathcal{O}\,\varphi := \neg\,\mathcal{P}\,\neg\varphi$

Permission in Standard Deontic Logic (SDL)

Deontic Logic: reasoning about norms (obligations, prohibitions, permissions)

- Introduced by Von Wright, 1951:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \vee \varphi) \mid (\varphi \rightarrow \varphi) \mid \mathcal{P}\,\varphi$$

- φ is **permitted**: $\mathcal{P}\,\varphi$
- φ is **forbidden**: $\mathcal{F}\,\varphi := \neg\,\mathcal{P}\,\varphi$
- φ is **obligatory**: $\mathcal{O}\,\varphi := \neg\,\mathcal{P}\,\neg\varphi$

- **Obligation** implies **permission**: $\mathcal{O}\,\varphi \rightarrow \mathcal{P}\,\varphi$
- Kripke Semantics

Paradoxes

- Free choice inference
- Ross' paradox

Paradoxes

- Free choice inference
  1. You may have coffee or tea
  2. Therefore, you may have coffee
  3. **and**, therefore, you may have tea

  $\mathcal{P}(\phi \vee \psi) \rightarrow \mathcal{P}(\phi) \wedge \mathcal{P}(\psi)$

- Ross' paradox

Paradoxes

- Free choice inference
  1. You may have coffee or tea
  2. Therefore, you may have coffee
  3. **and**, therefore, you may have tea

  $\mathcal{P}(\phi \vee \psi) \rightarrow \mathcal{P}(\phi) \wedge \mathcal{P}(\psi)$
  BUT we derive formulas such as: $\mathcal{O}(\phi) \rightarrow \mathcal{O}(\phi \wedge \psi)$

- Ross' paradox

Paradoxes

- Free choice inference
  1. You may have coffee or tea
  2. Therefore, you may have coffee
  3. **and**, therefore, you may have tea

  $\mathcal{P}(\phi \vee \psi) \rightarrow \mathcal{P}(\phi) \wedge \mathcal{P}(\psi)$
  BUT we derive formulas such as: $\mathcal{O}(\phi) \rightarrow \mathcal{O}(\phi \wedge \psi)$

- Ross' paradox

  You may read the letter. $\mathcal{P}(\phi)$

  BUT by monotonicity of permission (since $\phi \rightarrow \phi \vee \psi$):

  You may read the letter or burn it. $\mathcal{P}(\phi \vee \psi)$

Types of permission

Types of permission

- Weak permission: ‘it is not forbidden to dance silly in public’
- Strong permission: ‘it is permitted to cross the street at the traffic light’

Types of permission

- Weak permission: ‘it is not forbidden to dance silly in public’
- Strong permission: ‘it is permitted to cross the street at the traffic light’
- Unilateral permission: ‘an obligation to appear in court implies a permission to enter the courtroom’
- Bilateral permission: ‘a permission to have tea implies a permission to not have tea’

Types of permission

- Weak permission: ‘it is not forbidden to dance silly in public’
- Strong permission: ‘it is permitted to cross the street at the traffic light’
- Unilateral permission: ‘an obligation to appear in court implies a permission to enter the courtroom’
- Bilateral permission: ‘a permission to have tea implies a permission to not have tea’
- Right: ‘the right to vote’
- Exception: ‘It is permitted to smoke, only in that area’

What do we have?

- An ambiguous concept
- A faulty formalization

Outline

1. Permission in Deontic Logic
2. Mīmāṃsā Deontic
3. Mīmāṃsā Permission
4. Incorperating Preferences

The Mīmāṃsā school

Mīmāṃsā is one of the most important schools of Hindu philosophy

- ca 2500 years of deontic investigations
- Focus on the interpretation of the prescriptive portions of the *Vedas*, a book of commands such as: do not kill living beings
- Some of these commands seem contradictory
- These commands are interpreted with ***nyāyas*** (rules), in order to get rid of the contradictions

The Mīmāṃsā school

Mīmāṃsā is one of the most important schools of Hindu philosophy

- ca 2500 years of deontic investigations
- Focus on the interpretation of the prescriptive portions of the *Vedas*, a book of commands such as: do not kill living beings
- Some of these commands <u>seem</u> contradictory
- These commands are interpreted with ***nyāyas*** (rules), in order to get rid of the contradictions
- Some of these rules can be translated to Hilbert axioms!

Methodology

Goal: extract logical properties from Mīmāṃsā texts

Original texts

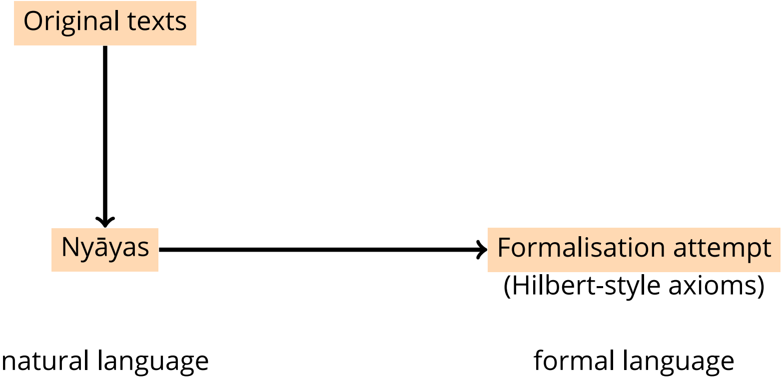
↓

Nyāyas

natural language

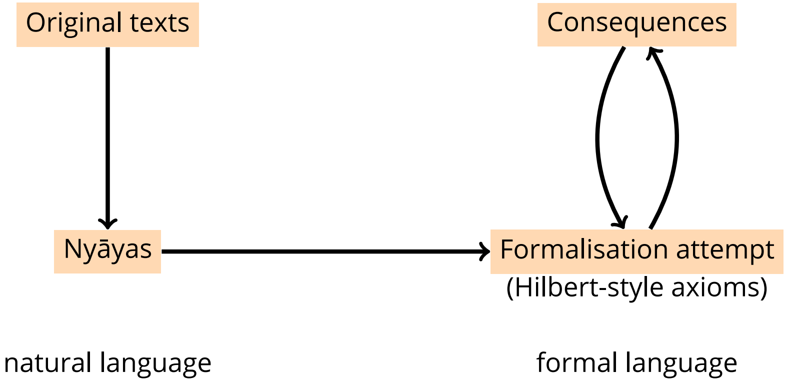
Methodology

Goal: extract logical properties from Mīmāṃsā texts



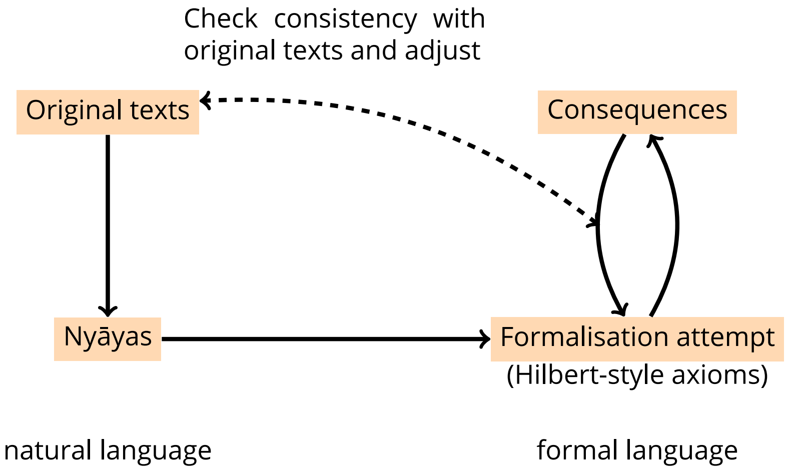
Methodology

Goal: extract logical properties from Mīmāṃsā texts



Methodology

Goal: extract logical properties from Mīmāṃsā texts



Mīmāṃsā obligation and prohibition

Mīmāṃsā obligation and prohibition

The base calculus:

- **Classical Logic**

Mīmāṃsā obligation and prohibition

The base calculus:

- **Classical Logic**
  - *Reductio ad absurdum* is admitted by Mīmāṃsā authors.

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

Mīmāṃsā obligation and prohibition

The base calculus:

- **Classical Logic**
  - *Reductio ad absurdum* is admitted by Mīmāṃsā authors.

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

- All operators are **dyadic**, e.g. $\mathcal{O}(\phi/\psi)$,
  Each command is uttered with regard to a specific eligible/responsible person (*adhikārin*) or to a specific situation.

Mīmāṃsā obligation and prohibition

The base calculus:

- **Classical Logic**
  - *Reductio ad absurdum* is admitted by Mīmāṃsā authors.

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

- All operators are **dyadic**, e.g. $\mathcal{O}(\phi/\psi)$,
  Each command is uttered with regard to a specific eligible/responsible person (*adhikārin*) or to a specific situation.
- **Not interdefinable**: fulfilling an obligation leads to a reward (or desired result), while trangressing a prohibition to a punishment. E.g. "It is forbidden to lie" $\neq$ "It is obligatory not to lie".

Resulting Logic

$$\phi ::= p \mid \phi \vee \phi \mid \neg\phi \mid \mathcal{O}(\phi/\psi) \mid \mathcal{F}(\phi/\psi) \mid \boxed{U}\phi$$

Ax1. $(\boxed{U}(\phi \to \psi) \wedge \mathcal{O}(\phi/\theta) \wedge \neg\boxed{U}\psi) \to \mathcal{O}(\psi/\theta)$

Ax2. $(\boxed{U}(\phi \to \psi) \wedge \mathcal{F}(\psi/\theta) \wedge \Diamond_U\phi) \to \mathcal{F}(\phi/\theta)$

Ax3. $\neg(X(\phi/\theta) \wedge X(\neg\phi/\theta))$ for $X \in \{\mathcal{O}, \mathcal{F}\}$

Ax4. $\neg(\mathcal{O}(\phi/\theta) \wedge \mathcal{F}(\phi/\theta))$

Ax5. $(\boxed{U}(\psi \leftrightarrow \theta) \wedge X(\phi/\psi)) \to X(\phi/\theta)$ for $X \in \{\mathcal{O}, \mathcal{F}\}$

Ax6. $(\Diamond_U(\phi \wedge \theta) \wedge \mathcal{O}(\phi/\top) \wedge \mathcal{O}(\theta/\top)) \to \mathcal{O}(\phi \wedge \theta/\top)$

S5 axioms for the global modality: $\boxed{U}$

Modelled with: Neighborhood Semantics $\langle W, N_{\mathcal{O}}, N_{\mathcal{F}}, V\rangle$.

Resulting Logic

$$\phi ::= p \mid \phi \vee \phi \mid \neg\phi \mid \mathcal{O}(\phi/\psi) \mid \mathcal{F}(\phi/\psi) \mid \boxed{U}\phi$$

Ax1. $(\boxed{U}(\phi \to \psi) \wedge \mathcal{O}(\phi/\theta) \wedge \neg\boxed{U}\psi) \to \mathcal{O}(\psi/\theta)$

Ax2. $(\boxed{U}(\phi \to \psi) \wedge \mathcal{F}(\psi/\theta) \wedge \Diamond_U\phi) \to \mathcal{F}(\phi/\theta)$

Ax3. $\neg(X(\phi/\theta) \wedge X(\neg\phi/\theta))$ for $X \in \{\mathcal{O}, \mathcal{F}\}$

Ax4. $\neg(\mathcal{O}(\phi/\theta) \wedge \mathcal{F}(\phi/\theta))$

Ax5. $(\boxed{U}(\psi \leftrightarrow \theta) \wedge X(\phi/\psi)) \to X(\phi/\theta)$ for $X \in \{\mathcal{O}, \mathcal{F}\}$

Ax6. $(\Diamond_U(\phi \wedge \theta) \wedge \mathcal{O}(\phi/\top) \wedge \mathcal{O}(\theta/\top)) \to \mathcal{O}(\phi \wedge \theta/\top)$

S5 axioms for the global modality: $\boxed{U}$

Modelled with: Neighborhood Semantics $\langle W, N_{\mathcal{O}}, N_{\mathcal{F}}, V\rangle$.

van Berkel, K., A. Ciabattoni, E. Freschi, F. Gulisano, and M. Olszewski. 2022. *Deontic Paradoxes in Mīmāṃsā Logics: There and Back Again.* JOLLI 2021.

What do we have?

1. A thought-out source of deontic investigations
2. Hilbert style formalization of obligation and prohibition
3. Without paradoxes!

Outline

1 Permission in Deontic Logic

2 Mīmāṃsā Deontic

3 Mīmāṃsā Permission

4 Incorperating Preferences

Permission in Mīmāṃsā (1/2)

- Permission are also dyadic, i.e. $\mathcal{P}(\phi/\psi)$.

Permission in Mīmāṃsā (1/2)

- Permission are also dyadic, i.e. $\mathcal{P}(\phi/\psi)$.
- Exceptions to general prohibitions or negative obligation
  - $\mathcal{P}(\phi/\psi) \rightarrow (\mathcal{F}(\phi/\top) \vee \mathcal{O}(\neg\phi/\top))$
  - $(\mathcal{P}(\phi/\psi) \wedge (\mathcal{F}(\phi/\theta) \vee \mathcal{O}(\neg\phi/\theta))) \rightarrow \boxed{\text{U}}(\psi \rightarrow \theta)$

  e.g.: The permission to eat after buying Soma implies the prohibition to eat (or the obligation not to eat) before it (*Tantravārttika* on 1.3.4).

Permission in Mīmāṃsā (1/2)

- Permission are also dyadic, i.e. $\mathcal{P}(\phi/\psi)$.
- Exceptions to general prohibitions or negative obligation
  - $\mathcal{P}(\phi/\psi) \rightarrow (\mathcal{F}(\phi/\top) \vee \mathcal{O}(\neg\phi/\top))$
  - $(\mathcal{P}(\phi/\psi) \wedge (\mathcal{F}(\phi/\theta) \vee \mathcal{O}(\neg\phi/\theta))) \rightarrow \boxed{\cup}(\psi \rightarrow \theta)$
- Commands entail possibility
  - $(\mathcal{O}(\phi/\psi) \vee \mathcal{F}(\phi/\psi)) \rightarrow \Diamond_{\cup}(\phi \wedge \psi) \wedge \neg\boxed{\cup}(\phi \wedge \psi)$
    The seeming prohibition “The fire is not to be kindled on the earth, nor in the sky, nor in heaven” cannot be taken as a prohibition, because fire cannot be kindled in the sky nor in heaven (see ŚBh on 1.2.5 and 1.2.18).

Permission in Mīmāṃsā (1/2)

- Permission are also dyadic, i.e. $\mathcal{P}(\phi/\psi)$.
- Exceptions to general prohibitions or negative obligation
  - $\mathcal{P}(\phi/\psi) \rightarrow (\mathcal{F}(\phi/\top) \vee \mathcal{O}(\neg\phi/\top))$
  - $(\mathcal{P}(\phi/\psi) \wedge (\mathcal{F}(\phi/\theta) \vee \mathcal{O}(\neg\phi/\theta))) \rightarrow \boxed{U}(\psi \rightarrow \theta)$
- Commands entail possibility
  - $(\mathcal{O}(\phi/\psi) \vee \mathcal{F}(\phi/\psi)) \rightarrow \Diamond_U(\phi \wedge \psi) \wedge \neg\boxed{U}(\phi \wedge \psi)$
    The seeming prohibition “The fire is not to be kindled on the earth, nor in the sky, nor in heaven” cannot be taken as a prohibition, because fire cannot be kindled in the sky nor in heaven (see ŚBh on 1.2.5 and 1.2.18).
  - $\mathcal{P}(\phi/\psi) \rightarrow \Diamond_U\phi \wedge \neg\boxed{U}\phi$

Permission in Mīmāṃsā (1/2)

- Permission are also dyadic, i.e. $\mathcal{P}(\phi/\psi)$.
- Exceptions to general prohibitions or negative obligation
  - $\mathcal{P}(\phi/\psi) \rightarrow (\mathcal{F}(\phi/\top) \vee \mathcal{O}(\neg\phi/\top))$
  - $(\mathcal{P}(\phi/\psi) \wedge (\mathcal{F}(\phi/\theta) \vee \mathcal{O}(\neg\phi/\theta))) \rightarrow \boxed{U}(\psi \rightarrow \theta)$
- Commands entail possibility
  - $(\mathcal{O}(\phi/\psi) \vee \mathcal{F}(\phi/\psi)) \rightarrow \Diamond_U(\phi \wedge \psi) \wedge \neg\boxed{U}(\phi \wedge \psi)$
  - $\mathcal{P}(\phi/\psi) \rightarrow \Diamond_U\phi \wedge \neg\boxed{U}\phi$
- Permissions are better-not permissions

Permission in Mīmāṃsā (1/2)

- Permission are also dyadic, i.e. $\mathcal{P}(\phi/\psi)$.
- Exceptions to general prohibitions or negative obligation
  - $\mathcal{P}(\phi/\psi) \rightarrow (\mathcal{F}(\phi/\top) \vee \mathcal{O}(\neg\phi/\top))$
  - $(\mathcal{P}(\phi/\psi) \wedge (\mathcal{F}(\phi/\theta) \vee \mathcal{O}(\neg\phi/\theta))) \rightarrow \boxed{\mathsf{U}}(\psi \rightarrow \theta)$
- Commands entail possibility
  - $(\mathcal{O}(\phi/\psi) \vee \mathcal{F}(\phi/\psi)) \rightarrow \Diamond_{\mathsf{U}}(\phi \wedge \psi) \wedge \neg\boxed{\mathsf{U}}(\phi \wedge \psi)$
  - $\mathcal{P}(\phi/\psi) \rightarrow \Diamond_{\mathsf{U}}\phi \wedge \neg\boxed{\mathsf{U}}\phi$
- Permissions are better-not permissions
  e.g: If one still refrains from eating meat, even though eating it is permitted, this is a meritorious act which leads one to the accumulation of good *karman*. (TV ad ŚBh 1.3.4)

Permission in Mīmāṃsā (1/2)

- Permission are also dyadic, i.e. $\mathcal{P}(\phi/\psi)$.
- Exceptions to general prohibitions or negative obligation
  - $\mathcal{P}(\phi/\psi) \rightarrow (\mathcal{F}(\phi/\top) \vee \mathcal{O}(\neg\phi/\top))$
  - $(\mathcal{P}(\phi/\psi) \wedge (\mathcal{F}(\phi/\theta) \vee \mathcal{O}(\neg\phi/\theta))) \rightarrow \boxed{U}(\psi \rightarrow \theta)$
- Commands entail possibility
  - $(\mathcal{O}(\phi/\psi) \vee \mathcal{F}(\phi/\psi)) \rightarrow \Diamond_{U}(\phi \wedge \psi) \wedge \neg\boxed{U}(\phi \wedge \psi)$
  - $\mathcal{P}(\phi/\psi) \rightarrow \Diamond_{U}\phi \wedge \neg\boxed{U}\phi$
- *Permissions are better-not permissions*

Permission in Mīmāṃsā (2/2)

Permission in Mīmāṃsā (2/2)

- ✗ Interdefinability: $\phi ::= p \mid \neg\phi \mid \phi \vee \phi \mid \mathcal{O}(\phi/\phi) \mid \mathcal{F}(\phi/\phi) \mid \mathcal{P}(\phi/\phi) \mid \boxed{\cup}\phi$

Permission in Mīmāṃsā (2/2)

- ✗ Interdefinability: $\phi ::= p \mid \neg\phi \mid \phi \vee \phi \mid \mathcal{O}(\phi/\phi) \mid \mathcal{F}(\phi/\phi) \mid \mathcal{P}(\phi/\phi) \mid \boxdot\phi$
- ✗ Obligation implies permission
  a) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{F}(\phi/\psi))$
  b) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\phi/\psi))$
  c) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\neg\phi/\psi))$

Permission in Mīmāṃsā (2/2)

- ✗ Interdefinability: $\phi ::= p \mid \neg\phi \mid \phi \vee \phi \mid \mathcal{O}(\phi/\phi) \mid \mathcal{F}(\phi/\phi) \mid \mathcal{P}(\phi/\phi) \mid \boxed{\cup}\phi$
- ✗ Obligation implies permission
  a) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{F}(\phi/\psi))$
  b) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\phi/\psi))$
  c) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\neg\phi/\psi))$
- ✗ Monotonicity of permission
  $\boxed{\cup}(\phi \rightarrow \psi)$, $\mathcal{P}(\phi/\theta)$ implies $\mathcal{P}(\psi/\theta)$

Permission in Mīmāṃsā (2/2)

- ✗ Interdefinability: $\phi ::= p \mid \neg\phi \mid \phi \vee \phi \mid \mathcal{O}(\phi/\phi) \mid \mathcal{F}(\phi/\phi) \mid \mathcal{P}(\phi/\phi) \mid \boxed{\text{U}}\phi$
- ✗ Obligation implies permission
  a) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{F}(\phi/\psi))$
  b) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\phi/\psi))$
  c) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\neg\phi/\psi))$
- ✗ Monotonicity of permission
  $\boxed{\text{U}}(\phi \rightarrow \psi)$, $\mathcal{P}(\phi/\theta)$ implies $\mathcal{P}(\psi/\theta)$
  $\Rightarrow \mathcal{F}(\psi/\top) \vee \mathcal{O}(\neg\psi/\top)$

Resulting logic

$$\phi ::= p \mid \neg\phi \mid \phi \vee \phi \mid \mathcal{O}(\phi/\phi) \mid \mathcal{F}(\phi/\phi) \mid \mathcal{P}(\phi/\phi) \mid \boxed{U}\phi$$

Earlier axioms for $\mathcal{O}$ and $\mathcal{F}$, the S5 axioms for $\boxed{U}$

P1. $\mathcal{P}(\phi/\psi) \rightarrow (\mathcal{F}(\phi/\top) \vee \mathcal{O}(\neg\phi/\top))$

P2. (a) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{F}(\phi/\psi))$

(b) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\phi/\psi))$

(c) $\neg(\mathcal{P}(\phi/\psi) \wedge \mathcal{O}(\neg\phi/\psi))$

P3. $(\mathcal{O}(\phi/\psi) \vee \mathcal{F}(\phi/\psi)) \rightarrow \Diamond_U(\phi \wedge \psi) \wedge \neg\boxed{U}\phi$

P4. (a) $(\boxed{U}(\psi \leftrightarrow \theta) \wedge \mathcal{P}(\phi/\psi)) \rightarrow \mathcal{P}(\phi/\theta)$

(b) $(\boxed{U}(\phi \leftrightarrow \psi) \wedge \mathcal{P}(\phi/\theta)) \rightarrow \mathcal{P}(\psi/\theta)$

P5. $(\mathcal{P}(\phi/\psi) \wedge (\mathcal{F}(\phi/\theta) \vee \mathcal{O}(\neg\phi/\theta))) \rightarrow \boxed{U}(\psi \rightarrow \theta)$

Modelled with: Neighborhood Semantics $\langle W, N_{\mathcal{O}}, N_{\mathcal{F}}, N_{\mathcal{P}}, V\rangle$.

What do we have?

- Soundness and Completeness wrt neighborhood semantics

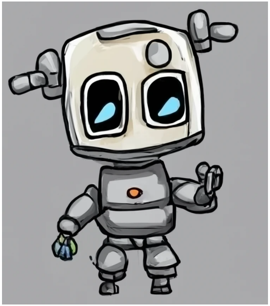
What do we have?

- Soundness and Completeness wrt neighborhood semantics
- Countermodels for the paradoxes
  - Free choice inference
  - Ross paradox
  - Privacy Act paradox

What do we have?

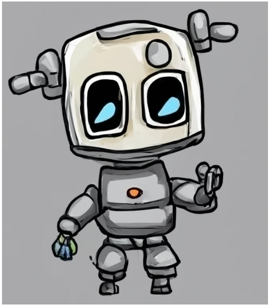
- Soundness and Completeness wrt neighborhood semantics
- Countermodels for the paradoxes
  - Free choice inference
  - Ross paradox
  - Privacy Act paradox
- Agata Ciabattoni, Josephine Dik, and Elisa Freschi. *Disambiguating Permissions: A Contribution from Mīmāṃsā.* DEON 2023.

BUT



Our robot still does not know what to do!

BUT



Our robot still does not know what to do!

Better-not permission?

Outline

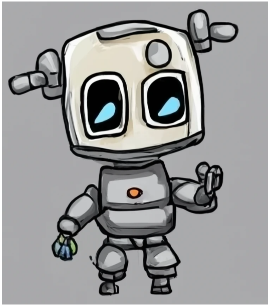
1 Permission in Deontic Logic

2 Mīmāṃsā Deontic

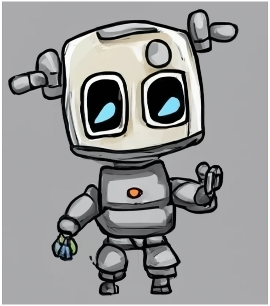
3 Mīmāṃsā Permission

4 Incorperating Preferences

Household Robot



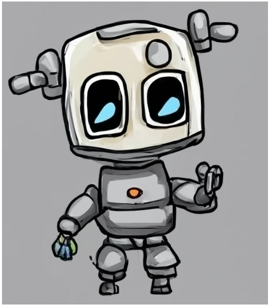
Household Robot



Humanoid robot working in a household setting

- Cleaning
- Assisting with entertainment
- Handle sharp objects

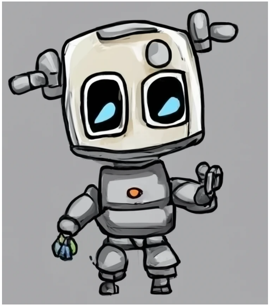
Household Robot



Humanoid robot working in a household setting

- Cleaning $\geq \neg$ Cleaning
- Assisting with entertainment
- Handle sharp objects

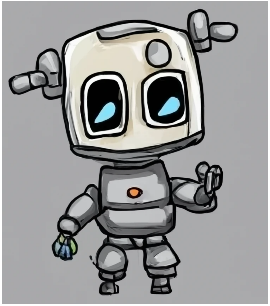
Household Robot



Humanoid robot working in a household setting

- Cleaning $\geq \neg$ Cleaning
- Assisting with entertainment $\sim \neg$ Assisting with entertainment
- Handle sharp objects

Household Robot



Humanoid robot working in a household setting

- Cleaning $\geq \neg$ Cleaning
- Assisting with entertainment $\sim$ $\neg$ Assisting with entertainment
- $\neg$ Handle sharp objects $\geq$ Handle sharp objects

The three permissions

- **Better-not permission**: it is permitted to hand me a sharp object when requested, but it is better not to.

The three permissions

- **Better-not permission**: it is permitted to hand me a sharp object when requested, but it is better not to.
- **Rather-so permission**: it is permitted to clean, it is even encouraged.

The three permissions

- **Better-not permission**: it is permitted to hand me a sharp object when requested, but it is better not to.
- **Rather-so permission**: it is permitted to clean, it is even encouraged.
- **Neutral permission**: it is permitted to assist with entertainment.

The three permissions

- **Better-not permission**: it is permitted to hand me a sharp object when requested, but it is better not to.
- **Rather-so permission**: it is permitted to clean, it is even encouraged.
- **Neutral permission**: it is permitted to assist with entertainment.

Goal: give a formal definition

Operator $\boxed{\preceq}$

Operator $\boxed{\preceq}$

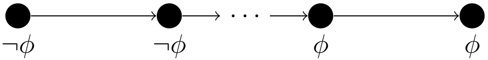
- $\boxed{\preceq}\phi$:= in all better scenarios, ϕ holds
  - $\boxed{\preceq}(\phi \to \psi) \to (\boxed{\preceq}\phi \to \boxed{\preceq}\psi)$
  - $\boxed{\preceq}\phi \to \boxed{\preceq}\boxed{\preceq}\phi$
  - $\boxed{\preceq}\phi \to \phi$
- $M = \langle W, N_\chi, \leq, V\rangle$ (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$)
  - where $\leq$ is transitive and reflexive
- $M, w \models \boxed{\preceq}\phi$ iff $\forall v\ w \leq v\ M, v \models \phi$

Operator $\boxed{\preceq}$

- $\boxed{\preceq}\phi$:= in all better scenarios, ϕ holds
  - $\boxed{\preceq}(\phi \rightarrow \psi) \rightarrow (\boxed{\preceq}\phi \rightarrow \boxed{\preceq}\psi)$
  - $\boxed{\preceq}\phi \rightarrow \boxed{\preceq}\boxed{\preceq}\phi$
  - $\boxed{\preceq}\phi \rightarrow \phi$
- $M = \langle W, N_\chi, \leq, V\rangle$ (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$)
  - where $\leq$ is transitive and reflexive
- $M, w \vDash \boxed{\preceq}\phi$ iff $\forall v\ w \leq v\ M, v \vDash \phi$
- 'ϕ is preferred over $\neg\phi$, in a context ψ' is translated to $\boxed{\text{U}}(\psi \rightarrow (\phi \rightarrow \boxed{\preceq}(\psi \rightarrow \phi)))$

Operator $\boxed{\preceq}$

- $\boxed{\preceq}\phi$:= in all better scenarios, ϕ holds
  - $\boxed{\preceq}(\phi \to \psi) \to (\boxed{\preceq}\phi \to \boxed{\preceq}\psi)$
  - $\boxed{\preceq}\phi \to \boxed{\preceq}\boxed{\preceq}\phi$
  - $\boxed{\preceq}\phi \to \phi$
- $M = \langle W, N_\chi, \leq, V\rangle$ (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$)
  - where $\leq$ is transitive and reflexive
- $M, w \vDash \boxed{\preceq}\phi$ iff $\forall v\ w \leq v\ M, v \vDash \phi$
- 'ϕ is preferred over $\neg\phi$, in a context ψ' is translated to $\boxed{U}(\psi \to (\phi \to \boxed{\preceq}(\psi \to \phi)))$



But wait!

But wait!

- **Example**: not engaging in the handling of sharp objects and breaking a law is not better than handling a sharp object and complying with all laws

But wait!

- **Example**: not engaging in the handling of sharp objects and breaking a law is not better than handling a sharp object and complying with all laws
- **Example:** the house being on fire and cleaning is not better than not cleaning in a house that is not on fire.

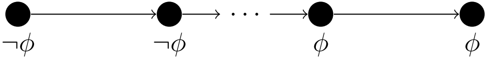
But wait!

- **Example**: not engaging in the handling of sharp objects and breaking a law is not better than handling a sharp object and complying with all laws
- **Example:** the house being on fire and cleaning is not better than not cleaning in a house that is not on fire.

$\Rightarrow$ Ceteris paribus preferences!
$\Rightarrow$ Thus we say ϕ is preferred over ψ, assuming a set of conditions Γ is agreed on.

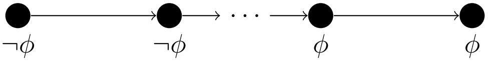
Operator $\boxed{\preceq}$

- $\boxed{\preceq}\phi$:= in all better scenarios, ϕ holds
  - $\boxed{\preceq}(\phi \rightarrow \psi) \rightarrow (\boxed{\preceq}\phi \rightarrow \boxed{\preceq}\psi)$
  - $\boxed{\preceq}\phi \rightarrow \boxed{\preceq}\boxed{\preceq}\phi$
  - $\boxed{\preceq}\phi \rightarrow \phi$
- $M = \langle W, N_\chi, \leq, V\rangle$, (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$),
  - where $\leq$ is transitive and reflexive
- $M, w \vDash \boxed{\preceq}\phi$ iff $\forall v\ w \leq v\ \ M, v \vDash \phi$
- 'ϕ is preferred over $\neg\phi$' is translated to $\boxed{U}(\psi \rightarrow (\phi \rightarrow \boxed{\preceq}(\psi \rightarrow \phi)))$



Operator $\boxed{\preceq}^{\Gamma}$

- $\boxed{\preceq}^{\Gamma}\phi$:= in all better scenarios that agree on Γ, ϕ holds
  - $\boxed{\preceq}^{\Gamma}(\phi \to \psi) \to (\boxed{\preceq}^{\Gamma}\phi \to \boxed{\preceq}^{\Gamma}\psi)$
  - $\boxed{\preceq}^{\Gamma}\phi \to \boxed{\preceq}^{\Gamma}\boxed{\preceq}^{\Gamma}\phi$
  - $\boxed{\preceq}^{\Gamma}\phi \to \phi$
- $w \equiv_{\Gamma} v$ iff for all $\gamma \in \Gamma$ ($M, w \vDash \gamma$ iff $M, v \vDash \gamma$)
- $M = \langle W, N_{\chi}, \leq, V \rangle$, (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$),
  - where $\leq$ is transitive and reflexive
- $M, w \vDash \boxed{\preceq}^{\Gamma}\phi$ iff $\forall v$ $w \leq v$ and $w \equiv_{\Gamma} v$, $M, v \vDash \phi$
- 'ϕ is preferred over $\neg\phi$' is translated to $\boxed{U}(\psi \to (\phi \to \boxed{\preceq}^{\Gamma}(\psi \to \phi)))$



What is Γ?

ϕ is better-not permitted $\Rightarrow$ the scenario with $\neg\phi$ true is better than the scenario with ϕ true, whenever they agree on:

What is Γ?

ϕ is better-not permitted $\Rightarrow$ the scenario with $\neg\phi$ true is better than the scenario with ϕ true, whenever they agree on:

- The deontic statements: $\mathcal{L}_{deon}$

What is Γ?

ϕ is better-not permitted $\Rightarrow$ the scenario with $\neg\phi$ true is better than the scenario with ϕ true, whenever they agree on:

- The deontic statements: $\mathcal{L}_{deon}$
- The truth of atomic propositions: *Atm*

What is Γ?

ϕ is better-not permitted $\Rightarrow$ the scenario with $\neg\phi$ true is better than the scenario with ϕ true, whenever they agree on:

- The deontic statements: $\mathcal{L}_{deon}$
- The truth of atomic propositions: *Atm*
- ... except for the atoms of compared action itself: $Atm(\phi)$

What is Γ?

ϕ is better-not permitted $\Rightarrow$ the scenario with $\neg\phi$ true is better than the scenario with ϕ true, whenever they agree on:

- The deontic statements: $\mathcal{L}_{deon}$
- The truth of atomic propositions: Atm
- ... except for the atoms of compared action itself: $Atm(\phi)$

$\Gamma := f(\phi) = \mathcal{L}_{deon} \cup Atm \setminus Atm(\phi)$

Formal definition

Formal definition

Better-not permission:

$$\mathcal{P}^{-}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \boxed{\cup}(\psi \rightarrow (\neg\phi \rightarrow \boxed{\preceq}^{\ulcorner}(\psi \rightarrow \neg\phi)))$$

Formal definition

Better-not permission:

$$\mathcal{P}^{-}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \boxed{\cup}(\psi \rightarrow (\neg\phi \rightarrow \boxed{\preceq}^{\ulcorner}(\psi \rightarrow \neg\phi)))$$

Rather-so permission:

$$\mathcal{P}^{+}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \boxed{\cup}(\psi \rightarrow (\phi \rightarrow \boxed{\preceq}^{\ulcorner}(\psi \rightarrow \phi)))$$

Formal definition

Better-not permission:

$$\mathcal{P}^{-}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \boxed{\cup}(\psi \rightarrow (\neg\phi \rightarrow \boxed{\preceq}^{\Gamma}(\psi \rightarrow \neg\phi)))$$

Rather-so permission:

$$\mathcal{P}^{+}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \boxed{\cup}(\psi \rightarrow (\phi \rightarrow \boxed{\preceq}^{\Gamma}(\psi \rightarrow \phi)))$$

Neutral permission:

$$\mathcal{P}^{0}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \neg\mathcal{P}^{+}(\phi/\psi) \wedge \neg\mathcal{P}^{-}(\phi/\psi)$$

Formal definition

Better-not permission:

$$\mathcal{P}^{-}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \boxed{\cup}(\psi \to (\neg\phi \to \boxed{\preceq}^{\Gamma}(\psi \to \neg\phi)))$$

Rather-so permission:

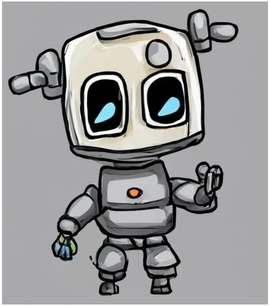
$$\mathcal{P}^{+}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \boxed{\cup}(\psi \to (\phi \to \boxed{\preceq}^{\Gamma}(\psi \to \phi)))$$

Neutral permission:

$$\mathcal{P}^{0}(\phi/\psi) ::= \mathcal{P}(\phi/\psi) \wedge \neg\mathcal{P}^{+}(\phi/\psi) \wedge \neg\mathcal{P}^{-}(\phi/\psi)$$

where $\Gamma := f(\phi) = \mathcal{L}_{deon} \cup Atm \setminus Atm(\phi)$

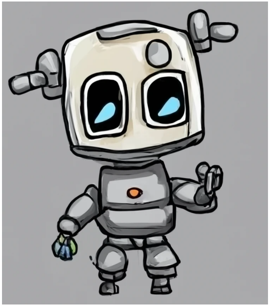
Household Robot



Humanoid robot working in a household setting.

- Cleaning $\geq \neg$ Cleaning
- Assisting in entertainment $\sim \neg$ Assisting in entertainment
- $\neg$ Handle sharp objects $\geq$ Handle sharp objects

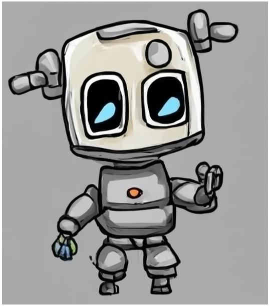
Household Robot



Humanoid robot working in a household setting.

- $cln \geq \neg cln$
- $ent \sim \neg ent$
- $\neg sharp \geq sharp$

Household Robot



Humanoid robot working in a household setting.

- $cln \geq \neg cln$ - $\mathcal{P}^{+}(cln/req)$
- $ent \sim \neg ent$ - $\mathcal{P}^{0}(ent/req)$
- $\neg sharp \geq sharp$ - $\mathcal{P}^{-}(sharp/req)$

Model

$M \models \mathcal{P}^{-}(sharp/req)$

Model

$M \vDash \mathcal{P}^{-}(sharp/req)$

$M = \langle W, N_\chi, \leq, V\rangle$ (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$, where

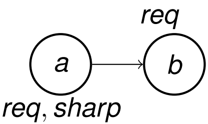
- $W = \{a, b\}$
- $N_{\mathcal{P}}(w) = \{(\|sharp\|, \|req\|)\}$ for all $w \in W$.

Model

$M \vDash \mathcal{P}^{-}(sharp/req)$

$M = \langle W, N_\chi, \leq, V \rangle$ (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$, where

- $W = \{a, b\}$
- $N_{\mathcal{P}}(w) = \{(\|sharp\|, \|req\|)\}$ for all $w \in W$.
- $V(req) = \{a, b\}$, $V(sharp) = \{a\}$
- Atm:= $\{sharp, req\}$
- $f(sharp) = \mathcal{L}_{deon} \cup \{req\}$

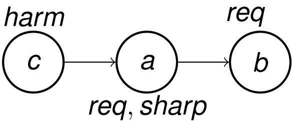


Model

$M \models \mathcal{P}^{-}(sharp/req)$

$M = \langle W, N_{\chi}, \leq, V \rangle$ (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$, where

- $W = \{a, b, c\}$
- $N_{\mathcal{P}}(w) = \{(\|sharp\|, \|req\|)\}$ for all $w \in W$.
- $V(req) = \{a, b\}$, $V(sharp) = \{a\}$, $V(harm) = \{c\}$
- Atm:= $\{sharp, req, harm\}$
- $f(sharp) = \mathcal{L}_{deon} \cup \{req, harm\}$

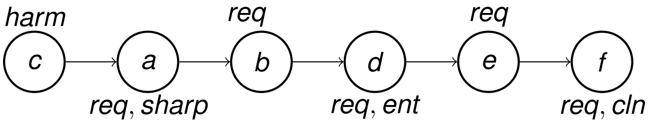


Model

$M \models \mathcal{P}^{-}(sharp/req) \wedge \mathcal{P}^{+}(cln/req) \wedge \mathcal{P}^{0}(ent/req)$

$M = \langle W, N_\chi, \leq, V \rangle$ (for $\chi \in \{\mathcal{O}, \mathcal{P}, \mathcal{F}\}$, where

- $W = \{a, b, c, d, e, f\}$
- $N_{\mathcal{P}}(w) = \{(\|sharp\|, \|req\|), (\|cln\|, \|req\|), (\|ent\|, \|req\|)\}$ for all $w \in W$.
- $V(req) = \{a, b, d, e, f\}$, $V(sharp) = \{a\}$, $V(harm) = \{c\}$,$V(ent) = \{d\}$, $V(cln) = \{f\}$
- Atm:= $\{sharp, req, harm, cln, ent\}$



Concluding remarks

Conclusion:

- 2500 years of deontic investigation led to very thought-out and inspiring definitions
  - Disambiguation of permission solving the paradoxes
  - Preference notion within permission

Concluding remarks

Conclusion:

- 2500 years of deontic investigation led to very thought-out and inspiring definitions
  - Disambiguation of permission solving the paradoxes
  - Preference notion within permission

Future work:

- Investigate which deontic logics to use for **implementation** of the three permissions
- A decision-making algorithm based on these permissions

Concluding remarks

Conclusion:

- 2500 years of deontic investigation led to very thought-out and inspiring definitions
  - Disambiguation of permission solving the paradoxes
  - Preference notion within permission

Future work:

- Investigate which deontic logics to use for **implementation** of the three permissions
- A decision-making algorithm based on these permissions

Long-term future work:

- Take inspiration from Mīmāṃsā deontic and apply to AI