Non-Monotonic Logic

Marco Degano

Philosophical Logic 2025
1 December 2025

Readings

Suggested:

- Frank Veltman, lecture notes on counterfactuals (sec. 5). `https://staff.fnwi.uva.nl/f.j.m.m.veltman/papers/Notes_Counterfactuals.pdf`
- S. Kraus, D. Lehmann & M. Magidor (1990), *Nonmonotonic Reasoning, Preferential Models and Cumulative Logics*.

Plan

1. Defeasible Reasoning
2. Cumulative Consequence Relation
3. Preferential Models
4. Applications

Outline

Defeasible reasoning

Birds fly.

- We often use **generalizations** that are *rationally compelling* but not deductively valid.
- We are talking about what *normally* or *typically* happens, not about exceptionless laws.
- Reasoning is **defeasible** when conclusions may have to be withdrawn in the light of further information, even if we keep our original premises.

Broad cases of defeasible reasoning

- **Everyday decision-making:** Lights are off in a café, so you assume it's closed. Then you see people inside and an "Open" sign: you keep "lights looked off from outside", but give up "the café is closed".
- **Social reasoning:** Your friend normally replies within minutes. This time they don't answer for hours, so you think they're upset. Then you learn they were on a long flight: you keep that they usually reply quickly, but drop "they're upset with me".
- **Default categorization (Tweety):** From "Tweety is a bird" we infer "Tweety flies". Learning "Tweety is a penguin", we keep that Tweety is a bird and that birds normally fly, but reject "Tweety flies".

Monotonicity: formal notion

Let $\models$ be a (single-conclusion) consequence relation between *sets* of formulas and formulas.

$\models$ is monotonic if for all sets Γ, Δ and all φ:

$$\text{if } \Gamma \models \varphi, \quad \text{then for every } \Delta \supseteq \Gamma, \ \Delta \models \varphi$$

Adding premises never invalidates an earlier consequence.

Classical consequence $\models$ (and standard proof systems for classical logic) satisfy monotonicity.

A consequence relation $\mid\sim$ is non-monotonic if there exist $\Gamma \subseteq \Delta$ and φ such that

$$\Gamma \mid\sim \varphi \quad \text{but} \quad \Delta \not\mid\sim \varphi$$

New information can *defeat* previous conclusions.

Non-monotonic reasoning in practice

We use a non-monotonic consequence symbol $\mid\!\sim$ (read $\alpha \mid\!\sim \beta :=$ if α, then normally β):

$$\text{Bird}(x) \mid\!\sim \text{Flies}(x)$$

- From the knowledge base $K = \{\text{Bird}(\text{Tweety})\}$ we may infer

$$K \mid\!\sim \text{Flies}(\text{Tweety}).$$

- After adding Penguin(Tweety), and a more specific default

$$\text{Penguin}(x) \mid\!\sim \neg\text{Flies}(x),$$

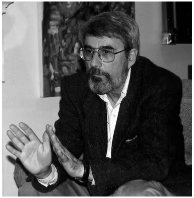
  the enlarged $K' = K \cup \{\text{Penguin}(\text{Tweety})\}$ no longer supports Flies(Tweety). Instead we get

$$K' \mid\!\sim \neg\text{Flies}(\text{Tweety}).$$

- This is exactly a failure of monotonicity.

From Aristotle to AI

John McCarthy
(1927–2011)



Raymond Reiter
(1939–2002)

- Aristotle already distinguished strict demonstration from more tentative, practical reasoning based on generalizations.
- In modern logic, **non-monotonic logic** is the study of formal systems intended to capture defeasible reasoning patterns.
- In AI and knowledge representation, many formalisms were proposed:
  - **Negation as failure** in logic programming.
  - **Circumscription** (McCarthy).
  - **Default logic** (Reiter).
  - ...
- These different formalisms all induce some *non-monotonic consequence relation* $\mid\!\sim$ on formulas.

Today's focus: the KLM perspective

Sarit Kraus

Daniel Lehmann

Menachem Magidor

Kraus, Lehmann & Magidor (KLM):

- Treat $\mid\!\sim$ itself as primitive: $\varphi \mid\!\sim \psi$ reads

  "from φ, it *normally* follows ψ."

- *Which structural rules should a reasonable non-monotonic consequence relation satisfy?*
  - **C** (cumulative relations).
  - **CL** (cumulative with Loop).
  - **P** (preferential relations).
  - **R** (rational relations, later work).
  - and some further, stronger systems.
- Prove *representation theorems*: each such family corresponds to a natural class of models (with preferences between worlds/states).

Today: focus on **cumulative and preferential relations** and system $\mathbf{C}$ and $\mathbf{P}$.

Outline

Cumulative consequence $\mathbf{C}$

A consequence relation $\mid\!\sim$ is **cumulative** iff it satisfies the rules:

1. **Reflexivity:** $\varphi \mid\!\sim \varphi$
2. **Left logical equivalence:** if $\varphi \models \psi$ and $\psi \models \varphi$, and $\varphi \mid\!\sim \chi$, then $\psi \mid\!\sim \chi$.
3. **Right weakening:** if $\varphi \models \psi$ and $\chi \mid\!\sim \varphi$, then $\chi \mid\!\sim \psi$.
4. **Cut:** if $\varphi \wedge \psi \mid\!\sim \chi$ and $\varphi \mid\!\sim \psi$, then $\varphi \mid\!\sim \chi$.
5. **Cautious monotonicity:** if $\varphi \mid\!\sim \psi$ and $\varphi \mid\!\sim \chi$, then $\varphi \wedge \psi \mid\!\sim \chi$.

- System $\mathbf{C}$ is meant to be the *minimal* structural core of reasonable non-monotonic consequence.
- $\mid\!\sim$ is closed under classical equivalence and consequence.
- Plausible conclusions can be “re-used” (Cut).
- Learning something you *already* inferred as plausible never harms (CMon).

Some derived rules (on blackboard)

In system $\mathbf{C}$ we can derive, among others, the following rules.

- **Equivalence**

$$\frac{\alpha \mathrel{|\!\sim} \beta \quad \beta \mathrel{|\!\sim} \alpha \quad \alpha \mathrel{|\!\sim} \gamma}{\beta \mathrel{|\!\sim} \gamma}$$

  If α and β are plausible consequences of each other, then they have the same plausible consequences.

- **Another rule**

$$\frac{\alpha \vee \beta \mathrel{|\!\sim} \alpha \quad \alpha \mathrel{|\!\sim} \gamma}{\alpha \vee \beta \mathrel{|\!\sim} \gamma}$$

  If from $\alpha \vee \beta$ we plausibly recover α, and from α we plausibly get γ, then already $\alpha \vee \beta$ plausibly entails γ.

Cumulative models

We give a semantics for system $\mathbf{C}$. Fix a set W of worlds.

A **cumulative model**

$$\mathcal{M}_C = \langle S, \prec, V \rangle$$

has:

- A set S of states (sets of worlds, possible "epistemic states" of an agent).
- A labelling function $\ell : S \to \mathcal{P}(W) \setminus \{\emptyset\}$:

  $\ell(s)$ is the set of worlds compatible with state s.
- A binary relation $\prec$ on S expressing *preference / normality*:

  $s' \prec s$ = "s' is more normal (preferred) than s".
- A valuation V assigning truth values to atoms at worlds:

  $V(w, p) \in \{0, 1\}$ for each world $w \in W$ and atomic p.

  Classical satisfaction $\mathcal{M}_C, w \models \alpha$ is defined in the usual way.

Cumulative models

Definition (State satisfaction)

Let $\mathcal{M}_C = \langle S, \ell, \prec, V \rangle$. For a state $s \in S$ and formula α:

$$s \models \alpha \quad \text{iff} \quad \forall w \in \ell(s)\ (\mathcal{M}_C, w \models \alpha).$$

So s satisfies α iff all its worlds do.

$$[\![\alpha]\!]^{\mathcal{M}_C} = \{s \in S \mid s \models \alpha\}$$

for the set of states that satisfy α.

Cumulative models

Definition (Smoothness for states)

A subset $A \subseteq S$ is **smooth** (with respect to $\prec$) iff for every $s \in A$:

- either s is $\prec$-minimal in A,
- or there is some $\prec$-minimal $s' \in A$ with $s' \prec s$.

$\mathcal{M}_C$ is a **cumulative model** iff for every formula α, the set $[\![\alpha]\!]^{\mathcal{M}_C}$ is smooth.

Smoothness is the analogue of the “no infinite descent” / limit assumption, now formulated for sets of *states*.

Example: a simple cumulative model

	p	q
w_1	1	1
w_2	1	0
w_3	0	1

Let $W = \{w_1, w_2, w_3\}$ and let V be the valuation given by the table. Define a structure

$$\mathcal{M}_C = \langle S, \ell, \prec, V \rangle$$

by:

$$S = \{s_0, s_1\}, \qquad \ell(s_0) = \{w_1, w_2\}, \qquad \ell(s_1) = \{w_3\}, \qquad s_0 \prec s_1.$$

$$[\![p]\!]^{\mathcal{M}_C} = \{s_0\}, \qquad [\![q]\!]^{\mathcal{M}_C} = \{s_1\}, \qquad [\![\top]\!]^{\mathcal{M}_C} = \{s_0, s_1\}.$$

In each of these sets the $\prec$-minimal elements are well behaved (e.g. s_0 is the unique minimal element of $[\![\top]\!]^{\mathcal{M}_C}$), so $\mathcal{M}_C$ satisfies the smoothness condition and is therefore a cumulative model.

How to make smoothness fail here? Add $s_1 \prec s_0$.

Consequence in cumulative models

Given a cumulative model

$$\mathcal{M}_C = \langle S, \ell, \prec, V \rangle$$

we define a model-relative consequence relation $\mid\!\sim_{\mathcal{M}_C}$.

Definition (Cumulative consequence in $\mathcal{M}_C$)

$\alpha \mid\!\sim_{\mathcal{M}_C} \beta$ iff for every $\prec$-minimal $s \in [\![\alpha]\!]^{\mathcal{M}_C}$, $s \models \beta$.

- Collect all states where α holds: $[\![\alpha]\!]^{\mathcal{M}_C}$.
- Restrict to the *best* (most normal) α-states (the $\prec$-minimal ones).
- If in all these best α-states every compatible world satisfies β, then β is a *cumulative consequence* of α.

Example: consequence in a cumulative model

Take atoms b, f, r.

	b	f	r
w_1	1	1	0
w_2	1	0	1
w_3	0	0	0

Let $W = \{w_1, w_2, w_3\}$ and let V be the valuation given by the table. Define a cumulative model

$$\mathcal{M}_C = \langle S, \ell, \prec, V \rangle$$

by

$$S = \{s_1, s_2, s_3\}, \quad \ell(s_1) = \{w_1\},\ \ell(s_2) = \{w_2, w_3\},\ \ell(s_3) = \{w_3\}$$

and a preference relation

$$s_1 \prec s_2 \qquad s_1 \prec s_3$$

with no further $\prec$-links.

$b \mid\!\sim_{\mathcal{M}_C} f \qquad b \wedge r \mid\!\sim_{\mathcal{M}_C} f \qquad \neg f \not\mid\!\sim_{\mathcal{M}_C} \neg r$

Representation theorem for system $\mathbf{C}$

Theorem (KLM representation for $\mathbf{C}$)

A consequence relation $\mid\!\sim$ on formulas is **cumulative** iff there exists a cumulative model

$$\mathcal{M}_C = \langle S, \ell, \prec, V \rangle$$

such that, for all formulas α, β,

$$\alpha \mid\!\sim \beta \quad \text{iff} \quad \alpha \mid\!\sim_{\mathcal{M}_C} \beta$$

- The proof rules of system $\mathbf{C}$ exactly capture reasoning in cumulative models.
- Preferential models are a *special case*:
  - S is a set of states, each labelled by a *single* world,
  - $\prec$ is a strict partial order on S.

Outline

1. Defeasible Reasoning

2. Cumulative Consequence Relation

3. Preferential Models

4. Applications

The move to preferential models

- In a **cumulative model**, a state $s \in S$ is an *information state*: $\ell(s) \subseteq W$ is the set of worlds compatible with what is currently taken for granted.
- States can therefore be "coarse-grained": one state may keep several classical possibilities open at once.
- A **preferential model** is the special case where every state is a singleton: $\ell(s) = \{w\}$. We *collapse* states with worlds and order the worlds directly.
- This makes the normality ordering more concrete: we compare "how things might be" world by world, just as in Lewis-Stalnaker similarity semantics for counterfactuals.
- Trade-off:
  - we gain simplicity and a tighter fit with the counterfactual picture
  - but lose generality: not every cumulative consequence relation can be represented by an ordering on single worlds.

Example

Take atoms p, q and worlds: $w_1 : p \wedge q \qquad w_2 : p \wedge \neg q$

Consider states:

$$s_{\text{coarse}} := \{w_1, w_2\} \quad \text{(“we know } p\text{, but } q \text{ is still open”)}$$

$$s_1 := \{w_1\} \qquad s_2 := \{w_2\}$$

We might prefer the “generic” state where only p is settled:

$$s_{\text{coarse}} \prec s_1, \qquad s_{\text{coarse}} \prec s_2.$$

In a pure **preferential** model, we only see w_1 and w_2. There is no separate node for the information state

“p is known, q is undetermined”.

So we lose the ability to order and reason about *information states* as such. We only order ‘fully specified’ ways the world might be.

The cumulative model above is actually ‘representable’ by a preferential model. How? Can you think of a case which cannot be represented in preferential models?

Example

Take atoms p, q, r and worlds:

$$w_1 : p \wedge q \wedge r \qquad w_2 : \neg p \wedge q \wedge r \qquad w_3 : p \wedge \neg q \wedge \neg r$$

Consider states:

$$s_p := \{w_1\} \qquad s_q := \{w_2\} \qquad s_{\text{coarse}} := \{w_2, w_3\}$$

We order states by:

$$s_{\text{coarse}} \prec s_p, \qquad s_{\text{coarse}} \prec s_q$$

$$p \mathrel{|\!\sim} r \qquad q \mathrel{|\!\sim} r \qquad (p \vee q) \not\mathrel{|\!\sim} r$$

There is no way, just by ordering *worlds*, to make the "most normal $p \vee q$-situation" behave like our coarse state $\{w_2, w_3\}$ that mixes r and $\neg r$ while still treating p and q separately as r-supporting.

Preferential models

A **preferential model** is a triple

$$\mathcal{M}_P = \langle W, \prec, V \rangle$$

such that:

- $\prec$ is a strict partial order on W
- for every formula α, the truth set $[\![\alpha]\!]^{\mathcal{M}_P} = \{w \in W \mid w \models \alpha\}$ is **smooth** with respect to $\prec$.

Definition (Smoothness for worlds)

A subset $A \subseteq W$ is **smooth** (with respect to $\prec$) iff for every $w \in A$:

- either w is $\prec$-minimal in A,
- or there is some $\prec$-minimal $w' \in A$ with $w' \prec w$.

Preferential consequence

Definition (Preferential consequence)

Given a preferential model $\mathcal{M}_P = \langle W, \prec, V \rangle$, define:

$\alpha \mid\!\sim_{\mathcal{M}_P} \beta$ iff for every $\prec$-minimal $w \in [\![\alpha]\!]^{\mathcal{M}_P},\ w \models \beta$.

$$[\![\alpha]\!]^{\mathcal{M}_P} = \{w \in W \mid \mathcal{M}_P, w \models \alpha\}$$

So we:

- look at all α-worlds,
- pick the *most normal* ones (the $\prec$-minimal α-worlds),
- and require all of them to satisfy β.

Knowledge bases and preferential entailment

A (default) knowledge base K is a set of conditionals

$$\alpha \mid\sim \beta$$

Definition (Satisfaction of a knowledge base)

Let K be a set of conditionals $\alpha \mid\sim \beta$. A preferential model $\mathcal{M}_P = \langle W, \prec, V \rangle$ *satisfies* K iff for every $\alpha \mid\sim \beta \in K$ we have

$$\alpha \mid\sim_{\mathcal{M}_P} \beta$$

Definition (Preferential entailment)

Let K be a set of conditionals. We write

$$K \models_{\text{pref}} \alpha \mid\sim \beta$$

iff for every preferential model $\mathcal{M}_P$: if $\mathcal{M}_P$ satisfies K, then

$$\alpha \mid\sim_{\mathcal{M}_P} \beta$$

Example: a bird-penguin knowledge base

Atoms: b (bird), f (flies), p (penguin).

Consider the knowledge base

$$K = \{p \mid\!\sim b, p \mid\!\sim \neg f, b \mid\!\sim f\}$$

- Any preferential model $\mathcal{M}_P$ that satisfies K must make the most normal p-worlds *non-flying birds*.
- In all such models we also have:

$$p \land b \mid\!\sim_{\mathcal{M}_P} \neg f$$

so

$$K \models_{\text{pref}} p \land b \mid\!\sim \neg f$$

- $K \not\models_{\text{pref}} p \mid\!\sim f$ [countermodel on blackboard]
- $K \models_{\text{pref}} f \mid\!\sim \neg p$ [proof on blackboard]

System P (preferential consequence)

System P is system C plus one extra rule, **Or**.

1. **Reflexivity** $\varphi \mathrel{|\!\sim} \varphi$
2. **Left logical equivalence**
   If $\varphi \models \psi$ and $\psi \models \varphi$, and $\varphi \mathrel{|\!\sim} \chi$, then $\psi \mathrel{|\!\sim} \chi$.
3. **Right weakening**
   If $\varphi \models \psi$ and $\chi \mathrel{|\!\sim} \varphi$, then $\chi \mathrel{|\!\sim} \psi$.
4. **Cut**
   If $\varphi \mathrel{|\!\sim} \psi$ and $\varphi \wedge \psi \mathrel{|\!\sim} \chi$, then $\varphi \mathrel{|\!\sim} \chi$.
5. **Cautious monotonicity**
   If $\varphi \mathrel{|\!\sim} \psi$ and $\varphi \mathrel{|\!\sim} \chi$, then $\varphi \wedge \psi \mathrel{|\!\sim} \chi$.
6. **Or**
   If $\varphi \mathrel{|\!\sim} \chi$ and $\psi \mathrel{|\!\sim} \chi$, then $\varphi \vee \psi \mathrel{|\!\sim} \chi$.

Reading the Or rule

Or: if $\varphi \mathrel{|\!\sim} \chi$ and $\psi \mathrel{|\!\sim} \chi$, then $\varphi \vee \psi \mathrel{|\!\sim} \chi$.

If both φ and ψ individually are good enough reasons for χ, then so is their disjunction.

(1)
a. If John comes to the party, it will normally be great.
b. If Cathy comes to the party, it will normally be great.
c. So if either John or Cathy comes, it will normally be great.

Some derived rules in $\mathbf{P}$ (blackboard)

In $\mathbf{P}$ we can derive several rules:

S-rule

$$\frac{\varphi \wedge \psi \mathrel{|\!\sim} \chi}{\varphi \mathrel{|\!\sim} (\psi \supset \chi)}$$

Union

$$\frac{\varphi \mathrel{|\!\sim} \psi \quad \chi \mathrel{|\!\sim} \gamma}{\varphi \vee \chi \mathrel{|\!\sim} \psi \vee \gamma}$$

Representation theorem

Theorem (KLM representation for $\mathbf{P}$)

A consequence relation $\mid\sim$ on formulas satisfies all rules of system $\mathbf{P}$ iff there exists a preferential model

$$\mathcal{M}_P = \langle W, \prec, V \rangle$$

such that, for all formulas φ, ψ,

$$\varphi \mid\sim \psi \quad \text{iff} \quad \varphi \mid\sim_{\mathcal{M}_P} \psi$$

For a knowledge base K, the *closure* of K under the rules of $\mathbf{P}$ coincides with preferential entailment:

$$K \vdash_{\mathbf{P}} \varphi \mid\sim \psi \quad \text{iff} \quad K \models_{\text{pref}} \varphi \mid\sim \psi$$

So to know what K *preferentially entails*, we can also reason syntactically with $\mathbf{P}$ instead of quantifying over all preferential models.

Duplicate labels and the representation theorem

Fix a propositional language with atoms p, q and this valuation:

	p	q
w_0	0	0
w_1	1	0
w_2	1	1
w_3	1	1

We define a preferential model $\mathcal{M}_P = \langle W, \prec, V \rangle$ by:

$$W = \{w_0, w_1, w_2, w_3\}$$

$w_0 \prec w_2, \qquad w_1 \prec w_3$ and no other $\prec$-links.

- This is a perfectly good preferential model (smoothness holds, $\prec$ is a strict partial order).
- It defines a consequence relation $\mid\!\sim_W$ that satisfies all rules of system $\mathbf{P}$, so $\mid\!\sim_W$ is a preferential consequence relation.
- However, there is **no** preferential model with *unique labels* (i.e. with taking worlds-as-valuations identifying w_2 with w_3) that induces exactly the same consequence relation $\mid\!\sim_W$.

Representation theorems

The **representation theorem** is a global, *structural* result about *consequence relations*:

- Fix a proof system (e.g. $\mathbf{C}$ or $\mathbf{P}$).
- Consider all binary relations $\mid\sim$ on formulas.
- The theorem says:

$$\mid\sim \text{ satisfies the rules of the system} \quad \Longleftrightarrow \quad \mid\sim = \mid\sim_{\mathcal{M}} \text{ for some model } \mathcal{M}$$

- So it *classifies* which abstract consequence relations are exactly those induced by a given class of models.

Soundness and Completeness (for a fixed entailment notion) is a more familiar, *formula-level* result:

- Fix a semantics (e.g. preferential models) and a proof system (e.g. $\mathbf{P}$).
- For a knowledge base K and a conditional $\alpha \mid\sim \beta$:

$$K \vdash_{\mathbf{P}} \alpha \mid\sim \beta \quad \Longleftrightarrow \quad K \models_{\text{pref}} \alpha \mid\sim \beta$$

- This says: whatever is valid in *all* models is derivable, and vice versa.

An underivable rule

Notably, $\mathbf{P}$ does *not* validate the following rule:[1]

Rational Monotonicity:

$$\text{if } \varphi \mid\!\sim \chi \text{ and } \varphi \not\mid\!\sim \neg\psi, \ \text{ then } \varphi \wedge \psi \mid\!\sim \chi.$$

Is this a good rule for non-monotonic reasoning?

We have already encountered this rule before. How was it called and to what example was related?

Similarity analysis of counterfactuals: *Strengthening with a Possibility* axiom scheme and *The Verdi-Bizet-Satie* example.

[1] Adding this rule to system $\mathbf{P}$ yields system $\mathbf{R}$ (Lehmann & Magidor 1992). Semantically, $\mathbf{R}$ corresponds to ranked models: there is a total preorder $\leq$ on W (a ranking of worlds) whose strict part is $\prec$. Thus any two worlds are comparable in rank, i.e. for all $v, w \in W$ we have $v \leq w$ or $w \leq v$ (or both).

Preferential Models and Counterfactuals

We can make the connection with the similarity framework we introduced for counterfactuals, with the following additional assumptions:

- The limit assumption: $\prec$ is a well-founded partial order on W.
- **Absoluteness:** for every $u, w \in W : \prec_u = \prec_w$ [$\prec_w$ is independent of w]
- **Universality:** for every $w \in W$, $W_w = W$ [the ordering is on W]

Recall the original clause for counterfactuals.

$M \models (\varphi \rightsquigarrow \psi)$ iff for every world $w \in W$, $M, u \models \psi$ for every closest $[\![\varphi]\!]$-world u to w.

With Universality and Absoluteness, we can simplify it as follows:

$M \models (\varphi \rightsquigarrow \psi)$ iff $M, u \models \psi$ for every $\prec$-minimal $[\![\varphi]\!]$-world u.

And then we rewrite $M \models (\varphi \rightsquigarrow \psi)$ as $\varphi \mid\!\sim_{\mathcal{M}_P} \psi$.

Object language vs metalanguage

- The conditional $\rightsquigarrow$ (for counterfactuals) is an *object-language* connective:

$$\varphi \rightsquigarrow \psi$$

is a formula that can itself be embedded.

- The non-monotonic consequence symbol $\mid\!\sim$ is a *metalanguage* relation:

$$\varphi \mid\!\sim \psi$$

is not a formula of L, but a statement about L.

- In the KLM approach, the central object of study *is* this relation $\mid\!\sim$ and its structural properties.

Outline

The frame problem

Consider designing how a robot should reason about actions and change.

(2) a. If the daylight sensor is low, turn on the light.
 b. If the temperature is low, turn on the heating.

In classical logic we might have:

$$\text{DaylightLow} \rightarrow \text{LightOn}$$
$$\text{TempLow} \rightarrow \text{HeatingOn}$$

But what about persistence?

- After turning the light on, should the robot keep believing that the light is *still* on at the next time step?
- Writing explicit “frame axioms” saying that everything stays the same unless affected by an action quickly leads to an explosion of axioms.

A non-monotonic take on the frame problem

Idea: use *default persistence* rules instead of explicit frame axioms.

Introduce discrete time steps $t, t+1, \ldots$ and write F_t for "F holds at time t".

For each F (light on, heating on, ...) we have a default:

$$F_t \mid\!\sim F_{t+1}$$

"if F holds at time t, then *normally* F still holds at time $t+1$".

- This replaces a huge family of classical frame axioms by a small, uniform *schema* of non-monotonic persistence rules.
- The rules are **defeasible**:

  e.g. if we also know that the action at t is *TurnOffLight*, then the default $LightOn_t \mid\!\sim LightOn_{t+1}$ is *defeated*.
- Non-monotonicity is crucial: new information about actions can make the robot *retract* a previous persistence conclusion without changing the earlier facts.

Grice and implicatures

(3) A: Does John speak English?
B: Well, he knows the colours.

- From B's answer we *normally* infer that John does not speak English (or at least not very well): if B could honestly say "Yes", that would be more informative.
- This is a **conversational implicature**: a defeasible inference driven by the assumption that speakers respect conversational maxims.
- The inference is **non-monotonic**: it can be cancelled without contradiction, e.g.
  B: Well, he knows the colours. In fact, his English is pretty good.

Pronoun resolution as default reasoning

(4) John met Bill at the station. *He* greeted *him*.

- By default, we resolve pronouns in line with simple preferences (e.g. subject $\rightarrow$ "he", object $\rightarrow$ "him"):

$$\text{John} = \textit{he}, \quad \text{Bill} = \textit{him}.$$

- This preferred interpretation is a **default**: it reflects what *normally* happens, given the syntax and discourse structure.
- But it is **defeasible**. Additional material can force a different resolution, e.g.

(5) John met Bill at the station. *He* greeted *him*.
Then John greeted him as well.

- Now we are pushed to reinterpret the first sentence so that *he* = Bill, *him* = John.

Temporal anaphora and non-monotonicity

- In narrative discourse with simple past tense, there is a strong **default**: events are understood as occurring in the order in which they are mentioned (forward-moving timeline).
- This default is **defeasible**: world knowledge or discourse relations can override it.

(6) John fell. Mary pushed him.

- The default forward-reading would place John's falling *before* Mary's pushing.
- But our knowledge about causation and the more natural discourse relation forces a different ordering: Mary pushed John *before* he fell.
- Asher & Lascarides (2003) give a systematic non-monotonic account of such temporal and discourse inferences (and related phenomena like lexical disambiguation).

Exercise: Soundness of rules in $\mathbf{C}$

- Show that the rules of system $\mathbf{C}$ are sound with respect to cumulative and preferential models.
- Show that the **Or** rule is sound with respect to preferential models.
- Show that the **Or** rule is not valid in cumulative models.

Exercise: Equivalence relation

We define an *equivalence relation* on formulas by:

$$\alpha \sim \beta \quad :\Longleftrightarrow \quad \alpha \mid\!\sim \beta \text{ and } \beta \mid\!\sim \alpha$$

Assume $\mid\!\sim$ is a cumulative consequence relation (i.e., it satisfies the rules of System $\mathbf{C}$).

Show that:

$$\alpha \sim \beta \quad \text{iff} \quad \forall\gamma\,(\alpha \mid\!\sim \gamma \;\Leftrightarrow\; \beta \mid\!\sim \gamma)$$

Exercise: Underivable rules in $\mathbf{P}$

Check that the following rules *cannot* be derived in $\mathbf{P}$:

1. If $\varphi \mathrel{|\!\sim} (\psi \supset \chi)$, then $\varphi \wedge \psi \mathrel{|\!\sim} \chi$.
2. If $\varphi \vee \psi \mathrel{|\!\sim} \chi$, then $\varphi \mathrel{|\!\sim} \chi$ or $\psi \mathrel{|\!\sim} \chi$.

Exercise: The **And** rule

And
If $\varphi \mid\sim \psi$ and $\varphi \mid\sim \chi$, then $\varphi \mid\sim \psi \wedge \chi$.

- Show that the **And** rule is a derived rule of system $\mathbf{C}$.
- Show semantically that **And** is valid in cumulative models, without using the representation theorem for $\mathbf{C}$.

Exercise: rules for $\mathbf{P}$

Show that the following system, with **And** in place of **Cut**, is an equivalent axiomatization of system $\mathbf{P}$:

1. **Reflexivity** $\varphi \mid\sim \varphi$.
2. **Left logical equivalence**
   If $\varphi \models \psi$ and $\psi \models \varphi$, and $\varphi \mid\sim \chi$, then $\psi \mid\sim \chi$.
3. **Right weakening**
   If $\varphi \models \psi$ and $\chi \mid\sim \varphi$, then $\chi \mid\sim \psi$.
4. **And**
   If $\varphi \mid\sim \psi$ and $\varphi \mid\sim \chi$, then $\varphi \mid\sim \psi \wedge \chi$.
5. **Cautious monotonicity**
   If $\varphi \mid\sim \psi$ and $\varphi \mid\sim \chi$, then $\varphi \wedge \psi \mid\sim \chi$.
6. **Or**
   If $\varphi \mid\sim \chi$ and $\psi \mid\sim \chi$, then $\varphi \vee \psi \mid\sim \chi$.

*Show that system $\mathbf{C}$ with **And** in place of **Cut** is is strictly weaker than $\mathbf{C}$ (define a non-monotonic consequence relation satisfying **Ref**, **LLE**, **RW**, **CMon**, and **And**, but not **Cut**.)

Exercise: The Penguin Triangle

$$p = \text{“penguin”} \quad b = \text{“bird”} \quad f = \text{“flies”}$$

Suppose K contains:

1. $p \mid\sim b$ (penguins are normally birds)
2. $p \mid\sim \neg f$ (penguins normally do not fly)
3. $b \mid\sim f$ (birds normally fly)

Show semantically or by taking the closure of K under $\mathbf{P}$:

1. $b \mid\sim \neg p$
2. $b \vee p \mid\sim f$
3. $b \vee p \mid\sim \neg p$

and explain informally why these are acceptable or problematic.