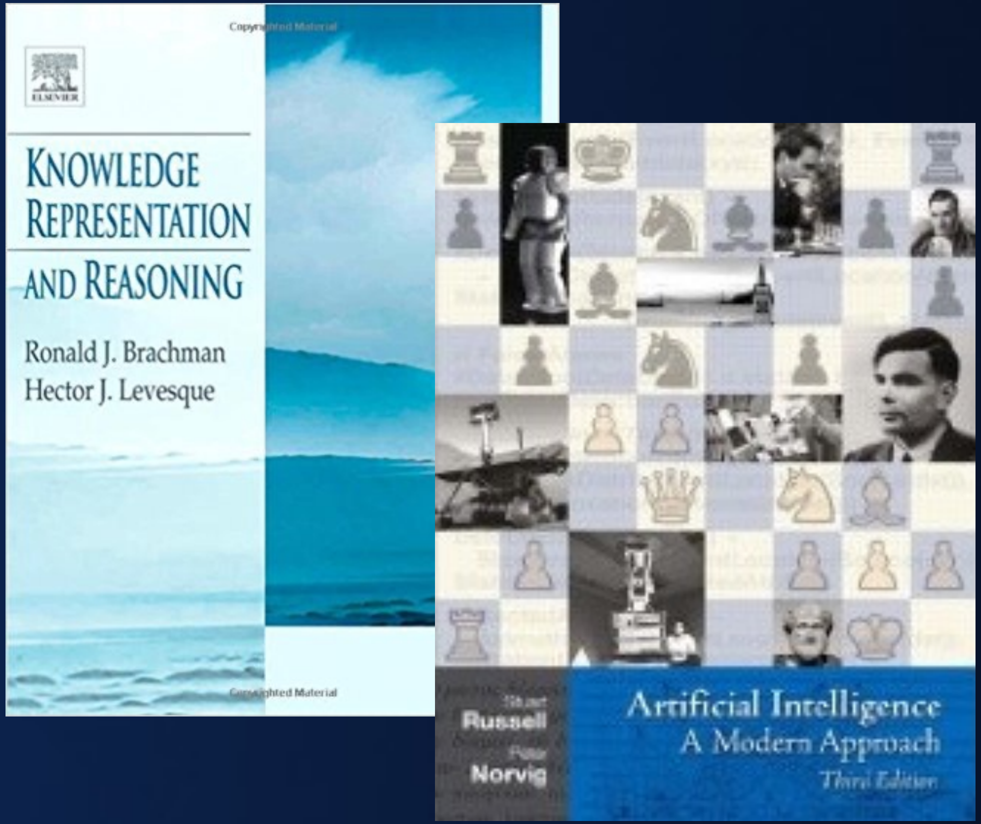


AI Fundamentals: Knowledge Representation and Reasoning

Maria Simi

Nonmonotonic reasoning

LESSON 3: CLOSED WORLD ASSUMPTION – CIRCUMSCRIPTION – DEFAULT LOGICS

Monotonicity of classical logic

Classical entailment is **monotonic.**

If $KB \vDash a$, then $KB \cup \{b\} \vDash a$ $\quad [KB \wedge b \vDash a]$

Failures of monotonicity are widespread in *commonsense reasoning*. It seems that humans often "jump to conclusions", when they think it is safe to do so (lacking information to the contrary). These conclusions are only "reasonable", given what you know, rather than **classically entailed**.

Most of the inference we do is *defeasible*: additional information, may lead to retract those tentative conclusions. Any time the set of beliefs does not grow monotonically when new evidence arrives, the monotonicity property is violated.

Including defeasible reasoning leads us to consider **nonsound inferences**.

Instances of nonmonotonic reasoning

Some common instances of nonmonotonic reasoning:

1. **Default reasoning**: reasonable assumptions unless evidence of the contrary
   - Car parked on the street; you assume it has four wheels even if you can see only two.
   - Birds fly, swan are white, bananas are yellow, tomatoes are red (prototypes).
2. **Persistence**: things stay the same, according to a principle of inertia, unless we know they change
3. **Economy of representation**: only true facts are stored, false facts are only assumed
4. **Reasoning about knowledge**: if you have $\neg Know(p)$ and you learn p ...
5. **Abductive reasoning**: most likely explanations to known facts.

Strictness of FOL universals

Universal rules, e.g. $\forall x\,(P(x) \Rightarrow Q(x))$

- express properties that apply to all instances
- all or nothing!

But most of what we learn about the world is in terms of **generics** rather than **universals**

Encyclopedia entries for ferry wheels, violins, turtles, wildflowers

E.g. "*Violins have four strings*" vs "*All violins have four strings*"

Properties are not strict for all instances, because

- genetic / manufacturing varieties
- borderline cases (early ferry wheels – toy violins)
- cases in exceptional circumstances etc.

Universal with exceptions

Listing exceptions is not a viable solution:

- *"All violins that are not E_1 or E_2 or ... have four strings"*.
  Exceptions usually are difficult to enumerate: *qualification problem.*
  Similarly, for general properties of individuals.
- Goal: be able to say a P is a Q in general, normally, but not necessarily. It is reasonable to conclude $Q(a)$, given $P(a)$, *unless there is a good reason not to.*
- This is what we call a **default** and **default reasoning** the tentative conclusion**.**
- Note: qualitative version (no numbers involved), probabilities are a different option to be examined later.

Approaches

There are three ways to approach the problem.

1. **Model theoretic formalizations** (CWA, Circumscription):
   - consist in a restriction to the possible interpretations, redefining the notion of entailment;
   - we can still have systems sound and complete *wrt* the new semantics.
2. **Proof theoretic formalizations** (Default logic, Autoepistemic logic)
   - A proof system with non-monotonic inference rules
   - Autoepistemic logic (under the heading “logics for knowledge and beliefs”)
3. **Systems supporting belief revision**
   - TMS, ATMS

Closed World Assumption (CWA)

- Reiter's observation:

  "There are usually many more negative facts than positive facts!"

  Example: airline flight database provides:

  $DirectConnect(cleveland, toronto)$ $DirectConnect(toronto, northBay)$,

  $DirectConnect(toronto, winnipeg)$ …

  but not: $\neg DirectConnect(cleveland, northBay)$ …

  The classical logical answer to $DirectConnect(cleveland, northBay)$: "I don't know"

- Under **Closed World Assumption (CWA)** only positive facts are stored, any other basic fact is assumed false.

  The answer to $DirectConnect(cleveland, northBay)$ under CWA : "No"

- CWA assumption is used in [deductive] databases and in logic programming with negation as failure.

Semantics of CWA

CWA:

Unless an atomic sentence is known to be true,

it can be assumed to be false.

CWA corresponds to a new version of entailment ($\models_c$):

Def CWA: $KB \models_c a$ *iff* $\mathrm{CWA}(KB) \models a$

where $\mathrm{CWA}(KB) = KB \cup \{\neg p \mid p$ ground atom and $KB \not\models p\}$

The set of assumed beliefs

$\mathrm{CWA}(KB)$ is the **completion** under CWA of KB

Note 1: CWA is a form of theory **completion**

Note 2: the CWA is nonmonotonic (computed as part of entailment)

Consistent and complete knowledge

- KB with consistent knowledge (satisfiable):

  For no α, $KB \vDash \alpha$ and $KB \vDash \neg\alpha$ *no contradiction*

Def Complete theory: For every α, $KB \vDash \alpha$ or $KB \vDash \neg\alpha$

- Normally, a KB has **incomplete knowledge:**

  Let $KB = \{p \vee q\}$ then $KB \vDash (p \vee q)$ but $KB \nvDash p$ and $KB \nvDash \neg p$

  Also, for any ground atom not mentioned in KB, $KB \nvDash r$ *and* $KB \nvDash \neg r$

- CWA can be seen as an assumption about **complete knowledge,** or a way to make a theory complete. If we consider a logic without quantifiers ...

  Theorem: For every α (within the language), $KB \vDash_c \alpha$ or $KB \vDash_c \neg\alpha$ **(1)**

  $\mathrm{CWA(KB)} \vDash \alpha$ or $\mathrm{CWA}(KB) \vDash \neg\alpha$

Consistency of CWA

Is $\mathrm{CWA}(KB)$ always consistent when KB is consistent? NO.

Problems with disjunctions:

Example. $KB = \{p \vee q\}$ $\quad \mathrm{CWA}(KB) = KB \cup \{\neg p, \neg q\}$ $\quad$ since $KB \not\models p$ and $KB \not\models q$

but $KB \cup \{\neg p, \neg q\} \models \neg(p \vee q)$ then $\mathrm{CWA}(KB)$ is inconsistent

Solution: restrict CWA to atoms that are "uncontroversial". p, q are controversial; r isn't.

CWA limited in such a way is called Generalized CWA (GCWA) and is a weaker form of completion than unrestricted CWA (the assumed beliefs are less)

GCWA. If $KB \models \{p \vee q_{1,} \dots \vee q_n\}$ and $KB \not\models p$ add $\neg p$

provided at least one ground positive literal q_i is entailed by KB.

Theorem: Consistency of CWA **(2)**

$\mathrm{CWA}(KB)$ is consistent *iff*

whenever $KB \models (q_1 \vee \dots \vee q_n)$ (where q_i are ground positive literals) then $KB \models q_i$ for some q_i

Horn KB consistency under CWA

Since it may be difficult to test the conditions of **Theorem 2** the following corollary, which restricts the application of CWA, is also of practical importance:

Corollary: If the KB is made of Horn clauses and it is consistent, then $\mathrm{CWA}(KB)$ is consistent.

Remember

Any FOL formula can be transformed into a set of clauses, preserving satisfiability.

Clause: a disjunctions of atomic formulas (positive and negative literals)

$\{l_1, l_2, \dots l_k\}$

Horn clause: *at most* one of the literals is positive. Horn clauses take one of two forms:

$\{p_1, \neg p_2, \dots \neg p_k\}$ or $\{\neg p_2, \dots \neg p_k\}$ where p's are positive literals

KB made of Horn clauses are the basis for rule-based systems and have other interesting properties.

Query evaluation

With CWA we can reduce queries (without quantifiers) to a set of atomic queries, by repeated applications of the following properties:

1. $KB \vDash_c (a \wedge b)$ iff $KB \vDash_c a$ and $KB \vDash_c b$
2. $KB \vDash_c \neg\neg a$ iff $KB \vDash_c a$
3. $KB \vDash_c \neg(a \vee b)$ iff $KB \vDash_c \neg a$ and $KB \vDash_c \neg b$
4. $KB \vDash_c (a \vee b)$ iff $KB \vDash_c a$ or $KB \vDash_c b$ *for KB completeness*
5. $KB \vDash_c \neg(a \wedge b)$ iff $KB \vDash_c \neg a$ or $KB \vDash_c \neg b$ *for KB completeness*

Example: $KB \vDash_c ((p \wedge q) \vee \neg(r \wedge \neg s))$

reduces to: either both $KB \vDash_c p$ and $KB \vDash_c q$ or $KB \vDash_c \neg r$ or $KB \vDash_c s$

Query evaluation (cont.)

If we further assume that CWA(KB) is consistent we get

6. $KB \vDash_c \neg a$ *iff* $KB \nvDash_c a$ (it cannot be $KB \vDash_c \neg a$ and $KB \vDash_c a$)

Example: $KB \vDash_c ((p \wedge q) \vee \neg(r \wedge \neg s))$

reduces to: either both $KB \vDash_c p$ and $KB \vDash_c q$ or $KB \nvDash_c r$ or $KB \vDash_c s$

Much more efficient than ordinary logic reasoning (e.g. no reasoning by cases).

In fact, we have restricted reasoning to a unique interpretation and instead of checking **validity/entailment** we check **truth** in that interpretation, a much more efficient task. We can also prove that:

Theorem: A KB with complete and consistent knowledge has a unique model.

Reasoning with models

1. Reasoning with sentences and entailment
2. Reasoning with an analogous representation of the world, a *model, a **vivid representation*** of the world.

To make this more intuitive, here is an analogy (see Brachman & Levesque):

Imagine the president of the United States standing directly beside the prime minister of Canada. It is observed that people have a hard time thinking about this scene without either imagining the president as being on the left or the prime minister as being on the left.

In a collection of sentences representing beliefs about the scene, we could easily leave out who is on the left. But in a model or diagram of the scene, we cannot represent the leaders as being beside each other without also committing to this and other visually salient properties of the scene.

Vivid knowlegde base

In a **vivid KB** we store this unique interpretation (a consistent and complete set of positive literals) and answer questions retrieving from it.

A vivid KB has the CWA builtin.

If positive atoms are stored as a table, deciding if $KB \models_c \alpha$ is like DB-retrieval.

Instead of reasoning with sentences we reason about an **analogical representation** of the world, with these properties:

- For each object of interest in the world, there is exactly one constant in KB+ that stands for that object;
- for each relationship of interest in the world, there is a corresponding predicate in the KB such that the relationship holds among certain objects in the world if and only if the predicate with the constants as arguments is an element of KB+.

Extension to quantifiers

The application of the theorem of consistency of CWA (Theorem 2) depends on the terms that we allow as part of the language.

Example: $KB = \{P(x) \vee Q(x), P(A), Q(B)\}$ and the only constants are A *and* B, then $\mathrm{CWA}(KB)$ is consistent. If we admit C, then $KB \models P(C) \vee Q(C)$, and $\mathrm{CWA}(KB)$ is not.

The **Domain Closure Assumption (DCA)** may be used to restrict the constants to those explicitly mentioned in the KB .

$\forall x\,.\,[x=c_1 \vee \ldots \vee x=c_n]$ where c_i are all the **finite** constants appearing in KB

Under this restriction quantifiers can be replaced by finite conjunctions and disjunctions.

The **Unique Names assumption** (**UNA**), can be used to deal with terms equality:

$(c_i \neq c_j)$, for $i \neq j$

With functions things get more complicated.

CWA in synthesis

- CWA is the assumption that atomic formulas not entailed by the KB are assumed to be false. This is a normal assumption in databases.

  Formally $KB \vDash_c a \;\; iff \;\; KB \cup \{\neg p \mid p$ ground atom and $KB \nvDash p\} \vDash a$
- The KB so augmented is **complete**: $\forall \alpha \;\; KB \vDash_{\mathrm{c}} \alpha$ or $KB \vDash_{\mathrm{c}} \neg \alpha$
- **Consistency** requires a more restricted formulation of the assumed belief

  GCWA. If KB $\vDash \{p \vee q_{1,} \ldots \vee q_n\}$ and $KB \nvDash p$ add $\neg p$ only if at least one ground literal q_i is entailed.
- Query processing can be reduced a combination of atomic queries.
- **Vivid** knowledge bases store the + part of a complete interpretation and make reasoning efficient (the augmentation reduces the possible models to one).

Predicate completion (only mention)

The CWA is too strong for many applications, We do not want to assume that **any** ground atom not provable from the KB is false: certain predicates are considered complete, others are not.

CWA wrt to a predicate P [set of predicates P]: the set of *assumed beliefs* is only for ground atoms in P [predicates in **P**].

A similar theory is **predicate completion** which consists in adding a set of completion axioms. If only $\mathrm{P}(A)$ and $\mathrm{P}(B)$ are in KB:

$$\forall x\,(x = A) \lor (x = B) \Rightarrow \mathrm{P}(x) \qquad \forall x\,(x = A) \lor (x = B) \Leftrightarrow \mathrm{P}(x) \;\; \textit{completion}$$

The theory accounting for **if and when** this leads to consistent augmentation is quite complex. See for example [Genesereth & Nilsson, ch 6.2].

A generalization of CWA and predicate completion is **circumscription**.

Circumscription: minimizing abnormality

- **Circumscription** can be seen as a more powerful and precise version of the CWA, working also for FOL, which was problematic and required further assumptions.
- The idea is to specify special **abnormality predicates** for dealing with exceptions.

For example, suppose we want to assert the default rule "*birds fly*":

$\forall x\, Bird(x) \land \neg \boldsymbol{Ab_f(x)} \Rightarrow Flies(x)$ *all normal birds fly*

We also have: $Bird(Tweety)$, $Bird(Chilly)$, $Chilly \neq Tweety, \neg Flies(Chilly)$,
We want to infer $Flies(Tweety)$, but $Tweety$ could satisfy Ab_1 in some model.
The idea is to **minimize abnormality.**

Circumscription: Given the unary predicate Ab, consider only interpretations where $\mathcal{I}[Ab_f]$ is **as small as possible**, relative to KB. It will include $Chilly$ but not $Tweety$.

Note: Circumscription is a semantic notion based on **minimal models** (a kind of **model preference logics**) due to MacCarthy 1980.

Minimal entailment

Let **P** be a set of unary abnormality predicates.

Let $\mathcal{I}_1$ and $\mathcal{I}_2$ two interpretations that agree on the values of constants and functions.

Ordering on interpretations:

$\mathcal{I}_1 < \mathcal{I}_2$ *iff* same domain and for every $P \in \mathbf{P}$ $\mathcal{I}_1[P] \subset \mathcal{I}_2[P]$ holds

Def Minimal entailment:

$KB \vDash_{\leq} \alpha$ *iff* α is true in $\mathcal{I}$ in **every minimal model** $\mathcal{I}$

Note 1: model ($\mathcal{I}[KB] = true$) and *minimal* (such that there is no other interpretations $\mathcal{I}' < \mathcal{I}$ such that $\mathcal{I}'[KB] = true$)

Note 2: α need not be true in all interpretations satisfying KB but **only in all those that minimize abnormalities.**

Example

Going back to the example:

$Bird(Tweety)$, $Bird(Chilly)$, $Chilly \neq Tweety$, $\neg Flies(Chilly)$

$Bird(x) \wedge \neg Ab(x) \Rightarrow Flies(x)$ *all normal birds fly*

Classically $KB \nvDash Flies(Tweety)$. However, $KB \vDash_{\leq} Flies(Tweety)$

The reason is this:

If $\mathcal{I}[KB] = true$ but $\mathcal{I}[Flies(Tweety)] = false$, then $\mathcal{I}[Ab(Tweety)] = true$.

So let $\mathcal{I}'$ be exactly $\mathcal{I}$ except that we remove the denotation of $Tweety$ from the interpretation of Ab. Then $\mathcal{I}' < \mathcal{I}$ (assuming **P** = $\{Ab\}$), and still $\mathcal{I}'[KB] = true$.

Thus, in the minimal models of the KB, Tweety is a normal bird:

$KB \vDash_{\leq} \neg Ab(tweety)$, and $KB \vDash_{\leq} Flies(tweety)$.

We cannot do the same for $Chilly$ and in fact $\neg Flies(Chilly)$.

Minimal models and CWA/CGWA

Circumscription need not produce a unique minimal interpretation

Suppose KB = {… , $Bird(C)$, $Bird(D)$, $(\neg Flies(C) \vee \neg Flies(D))$}

In this case there are two minimal models (one with $Flies(C)$ and one with $Flies(D)$) but we need to consider what is true **in all minimal models …**

$KB \models_{\leq} Flies(C) \vee Flies(D)$

$KB \not\models_{\leq} Flies(C)$ and $KB \not\models_{\leq} Flies(D)$

CWA would add the literal $\neg Ab(C)$, and by similar reasoning $\neg Ab(D)$, leading to an inconsistency.

GCWA would not conclude anything by default about C or D.

Circumscription is more cautious than CWA in the assumptions it makes about "controversial" individuals, like C and D.

Circumscription and quantified sentences

Circumscription works equally well with unnamed individuals. Suppose we add:

$$\exists x\, [Bird(x) \land (x \neq Tweety) \land (x \neq Chilly) \land InTree(x)]$$

We can conclude:

$$\exists x\, [Bird(x) \land (x \neq Tweety) \land (x \neq Chilly) \land InTree(x) \land Flies(x)]$$

In the minimal models there will be a single abnormal individual, *Chilly*.

If on the other end:

$$\neg Flies(Chilly) \land \exists x\, [Bird(x) \land (x \neq Chilly) \land (x \neq Tweety) \land \neg Flies(x)]$$

a minimal model will have exactly two abnormal individuals: *Chilly* and the other one that doesn't fly

Open issues with circumscription

Although the default assumptions made by circumscription are usually weaker than those of the CWA, there are cases where they appear too strong.

Suppose, for example, that we have the following KB:

$\forall x\,[Bird(x) \land \neg Ab(x) \Rightarrow Flies(x)]$

$Bird(tweety)$

$\forall x\,[Penguin(x) \Rightarrow (Bird(x) \land \neg Flies(x))]$

From this follows:

$\forall x\,[Penguin(x) \Rightarrow Ab(x)]$

Minimizing abnormalities leads to:

$KB \models_{\leq} \neg\exists x\, Ab(x)$

But also:

$KB \models_{\leq} \neg\exists x\, Penguin(x)$ i.e. there are no penguins, too strong

Partial fix

McCarthy's definition related to *predicate completion*. Distinguish between **P** (*variable* predicates) and **Q** (*fixed* predicates)

Ordering on interpretations is different for the two sets:

$\mathcal{I}_1 \leq \mathcal{I}_2$ iff same domain and

1. for every $P \in \mathbf{P}$ $\mathcal{I}_1[P] \subseteq \mathcal{I}_2[P]$ holds — **P** variable predicates
2. for every $Q \in \mathbf{Q}$ $\mathcal{I}_1[Q] = \mathcal{I}_2[Q]$ holds — **Q** fixed predicates

so only predicates in **P** are allowed to be minimized.

Previous example: **P** = $\{Ab\}$ and **Q** = {*Penguin*}; minimize Ab keeping $Penguin$ fixed. The only Ab are Penguins.

Problems:

- need to decide what to allow to vary: what about *Flies*?
- cannot conclude $Flies(Tweety)$ by default! $Tweety$ could be a penguin, thus *Ab* (only get default $(\neg Penguin(Tweety) \Rightarrow Flies(Tweety))$)

Default logic [Reiter]

We use rules to specify implicit beliefs. We distinguish:

- *explicit beliefs* = axioms
- *implicit beliefs* = theorems = least set closed under inference rules

Default logic KB uses two components: $KB = \langle F, D \rangle$

- F is a set of sentences (facts)
- D is a set of **default rules**: $\dfrac{\alpha : \beta}{\gamma}$

  read as "If you can infer α, and *it is consistent to assume* β, then infer γ"

  α : the prerequisite, β: the justification, γ: the conclusion

e.g. $\dfrac{Bird(tweety) : Flies(tweety)}{Flies(tweety)}$ also $\dfrac{Bird(x) : Flies(x)}{Flies(x)}$ (rule schema)

Default rules where $\beta = \gamma$ are called **normal defaults**

Extensions

Problem: how to characterize theorems/entailments

- cannot write a derivation, since do not know when to apply default rules
- no guarantee of unique set of theorems

Extensions: sets of sentences that are “reasonable” beliefs, given explicit facts and default rules

E is an **extension** of $\langle F, D \rangle$ iff for every sentence π, E satisfies the following:

$$\pi \in E \text{ iff } F \cup \Delta \vDash \pi \qquad \text{where } \Delta = \{\gamma \mid \frac{\alpha : \beta}{\gamma} \in D,\ \alpha \in E, \neg\beta \notin E\}$$

So, an extension E is the set of entailments of $F \cup \Delta$, where the Δ is a “suitable” set of assumptions given D. Note that α has to be in E, not in F. This has the effect of allowing the prerequisite to be believed as the result of other default assumptions.

Note that this definition is *not constructive*.

Example: single extension

Suppose KB is:

$$F = \{Bird(chilly), Bird(tweety), \neg Flies(chilly)\}$$

$$D = \{\frac{Bird(x) : Flies(x)}{Flies(\mathrm{x})}\}$$

then there is a unique extension, where $\Delta = \{Flies(tweety)\}$

- This is an extension since $Bird(tweety) \in E$ and $\neg Flies(tweety) \notin E$.
- No other extension, since $Flies(tweety)$ in any estension and no extension has $Flies(chilly)$.

If E is *inconsistent,* we can conclude anything we want.

Theorem: An extension of a default theory is inconsistent *iff* the original F is inconsistent.

In this case the extension is unique, but in general a default theory can have **multiple extensions**.

Example: multiple extensions

The **Nixon diamond**:

$F = \{Quacker(Nixon), Republican(Nixon)\}$

$D = \{Quacker(x) : Pacifist(x) \; / \; Pacifist(x),$

$Republican(x) : \neg Pacifist(x) \; / \; \neg Pacifist(x)\}$

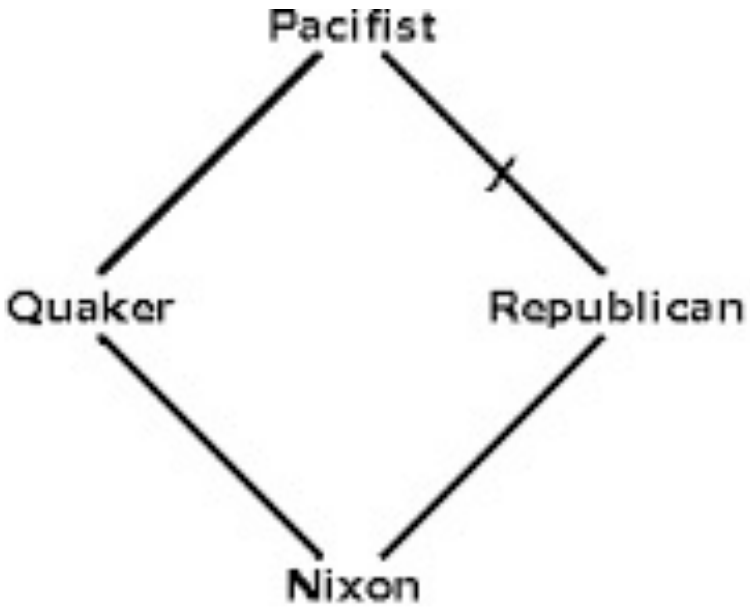
Two extensions:

E_1 has Δ = $Pacifist(Nixon)$

E_2 has Δ = $\neg Pacifist(Nixon)$

Which to believe? Two possible approaches:

1. **credulous**: choose an extension arbitrarily
2. **skeptical**: believe only what is common to all extensions



Properties

1. If a default theory has distinct extensions, they are **mutually inconsistent**.

   $F = \{A \vee B\}$ $D = \{:\neg A/\neg A, :\neg B/\neg B\}$ $E_1 = \{A \vee B, \neg A\}$ $E_2 = \{A \vee B, \neg B\}$

2. There are default theories with **no extensions.**

   Consider the default: $:A/\neg A.$ If $F = \{\,\}$ then $E = \{\,\}$

2. Any **normal** default theory has an extension.

3. Adding new normal default rules does not require the withdrawal of beliefs, even if adding new beliefs might. Normal default theories are **semi-nonmonotonic**.

Grounded extensions

We have a problem that leads to a more complex definition of extension.

Suppose $F = \{\ \}$ and $D = \{: p \,/\, p\}$

Then $E =$ entailments of $\{p\}$ is an extension since $p \in E$ and $\neg p \notin E$.

However, we have no good reason to believe p ! Only support for p is the default rule, which requires p itself as a prerequisite. So, the default should have no effect.

Desirable extension is only: $E =$ entailments of { }, i. e. all valid formulas.

A revision of the definition of extension is necessary. Reiter's definition:

Grounded extension: For any set S, let $\Gamma(S)$ be the **least set** containing F, closed under entailment and satisfying the default rules

$$\text{if } \alpha : \beta \,/\, \gamma \in D,\ \alpha \in \Gamma(S),\ \text{and } \neg\beta \notin S, \text{ then } \gamma \in \Gamma(S) \quad [\text{instead of } \neg\beta \notin \Delta(S)\ ]$$

A set E is an extension of $\langle F, D\rangle$ *iff* $\boldsymbol{E = \Gamma(E)}$**, i. e.** E is a fixed point of the Γ operator.

Conclusions

We have seen three ways of dealing with defaults and nonmonotonicity:

1. CWA: we try to complete the KB adding negative facts and use normal entailment. Nice computational properties for subsets of FOL.
2. Circumscription: tries to restrict the possible interpretations to the minimal ones, for certain predicates that we define as "*abnormal*".
3. Default logic: tries to characterize a new form of *tentative inference* through **default rules** and the notion of extensions.
4. Next, after talking about logic for knowledge and belief, we will add autoepistemic logic, which is a precursor of Answer Set Programming.

References

Knowledge Representation and Reasoning, Ronald Brachman and Hector Levesque, Elsevier, 2004 [Ch. 11]

Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach* (3rd edition). Pearson Education 2010 [Ch. 12]

Genesereth, M., and Nilsson, N., Logical Foundations of Artificial Intelligence, San Francisco: Morgan Kaufmann, 1987 [Ch. 6]