Strictness of FOL

To reason from $P(a)$ to $Q(a)$, need either

- facts about a itself
- universals, e.g. $\forall x(P(x) \supset Q(x))$
  - something that applies to all instances
  - all or nothing!

But most of what we learn about the world is in terms of <u>generics</u>

- e.g., encyclopedia entries for ferris wheels, violins, turtles, wildflowers

Properties are not strict for all instances, because

- genetic / manufacturing varieties
  - early ferris wheels
- borderline cases
  - toy violins
- imagined cases
  - flying turtles
- cases in exceptional circumstances
  - dried wildflowers
- ...

Generics vs. Universals

4 Violins have four strings

vs.

5 All violins have four strings

vs.

? All violins that are not E_1 or E_2 or ... have four strings

(exceptions usually cannot be enumerated)

Similarly, for general properties of individuals

Alexander the great: ruthlessness

Ecuador: exports

pneumonia: treatment

Goal: be able to say a P is a Q <u>in general</u>, but not necessarily

reasonable to conclude $Q(a)$ given $P(a)$
unless there is a good reason not to

Here: qualitative version (no numbers)

Varieties of defaults

General statements

- statistical: Most P's are Q's.
  - People living in Quebec speak French.
- normal: All normal P's are Q's.
  - Polar bears are white.
- prototypical: The prototypical P is a Q.
  - Owls hunt at night.

Representational

- conversational: Unless I tell you otherwise, a P is a Q.
  - default slot values in frames
  - disjointness in IS-A hierarchy (sometimes)
  - closed-world assumption (below)

Epistemic rationales

- familiarity: If a P was not a Q, you would know it.
  - an older brother
  - very unusual individual, situation or event
- group confidence: All known P's are Q's.
  - NP-hard problems unsolvable in poly time.

Persistence rationale

- inertia: A P is a Q if it used to be a Q.
  - colours of objects
  - locations of parked cars (for a while!)

Closed-world assumption

Reiter's observation:

There are usually many more -ve facts than +ve facts!

AirLine Example: flight guide provides

DirectConnect(cleveland,toronto)
DirectConnect(toronto,northBay)
DirectConnect(toronto,winnipeg) ...

but not: ¬DirectConnect(cleveland,northBay)

Conversational default, called CWA:

only +ve facts will be given, relative to some vocabulary

But note: KB $\not\models$ -ve facts

would have to answer: "I don't know"

Proposal: a new version of entailment

KB $\models_c \alpha$ iff KB $\cup$ *Negs* $\models \alpha$

where

$Negs = \{\neg p \mid p \text{ ground atomic and KB} \not\models p\}$

a common pattern: KB ' = KB $\cup \Delta$

Note: relation to negation as failure

Gives: KB $\models_c$ +ve facts and -ve facts

Properties of CWA

For every sentence α without quantifiers, either KB $\models_c \alpha$ or KB $\models_c \neg\alpha$ (or both)

Why? Inductive argument:

- immediately true for atomic sentences
- push $\neg$ in, e.g. KB $\models \neg\neg\alpha$ iff KB $\models \alpha$
- KB $\models (\alpha \wedge \beta)$ iff KB $\models \alpha$ and KB $\models \beta$
- Say KB $\not\models_c (\alpha \vee \beta)$.
  Then KB $\not\models_c \alpha$ and KB $\not\models_c \beta$.
  So by induction, KB $\models_c \neg\alpha$ and KB $\models_c \neg\beta$.
  Thus, KB $\models_c \neg(\alpha \vee \beta)$.

CWA is an assumption about <u>complete</u> knowledge

never any unknowns, relative to vocabulary

In general, a KB has <u>incomplete</u> knowledge,

e.g., if KB = $(p \vee q)$, then KB $\models (p \vee q)$, but
KB $\not\models p$, KB $\not\models \neg p$, KB $\not\models q$, and KB $\not\models \neg q$

But with CWA, always have:

If KB $\models_c (\alpha \vee \beta)$, then KB $\models_c \alpha$ or KB $\models_c \beta$

similar argument to above

Query evaluation

With CWA can reduce queries (without quantifiers) recursively to atomic case:

KB $\models_c (\alpha \wedge \beta)$ iff KB $\models_c \alpha$ and KB $\models_c \beta$

KB $\models_c (\alpha \vee \beta)$ iff KB $\models_c \alpha$ or KB $\models_c \beta$

KB $\models_c \neg(\alpha \wedge \beta)$ iff KB $\models_c \neg\alpha$ or KB $\models_c \neg\beta$

KB $\models_c \neg(\alpha \vee \beta)$ iff KB $\models_c \neg\alpha$ and KB $\models_c \neg\beta$

KB $\models_c \neg\neg\alpha$ iff KB $\models_c \alpha$

reduces to: KB $\models_c \lambda$, where λ is a literal

If KB $\cup$ *Negs* is consistent, get

KB $\models_c \neg\alpha$ iff KB $\not\models_c \alpha$

reduces to: KB $\models_c p$, where p is atomic

If atomic wffs stored as a table, deciding whether or not KB $\models_c \alpha$ is like DB-retrieval:

- reduce query to set of atomic queries
- solve atomic queries by table lookup

Different from ordinary logic reasoning

e.g. no reasoning by cases

see “vivid reasoning” (discussed later)

Consistency

If KB is set of atoms, then KB $\cup$ *Negs* is always consistent

Also works if KB has conjunctions and if KB has -ve disjunctions

If KB contains $(\neg p \vee \neg q)$. Add both $\neg p$, $\neg q$.

Problem when KB $\models (\alpha \vee \beta)$, but KB $\not\models \alpha$ and KB $\not\models \beta$

e.g. KB = $(p \vee q)$ *Negs* = $\{\neg p, \neg q\}$

so KB $\cup$ *Negs* is inconsistent

and for every α, KB $\models_c \alpha$!

Solution: only apply CWA to atoms that are "uncontroversial"

One approach: GCWA

Negs = $\{\neg p$ | If KB $\models (p \vee q_1 \vee ... \vee q_n)$ then KB $\models (q_1 \vee ... \vee q_n)\}$

When KB is consistent, get:

– KB $\cup$ *Negs* consistent

– everything derivable is also derivable by CWA

Quantifiers and Equality

So far, results do not extend to wffs with quantifiers

- can have KB $\not\models_c \forall x.\alpha$ and KB $\not\models_c \neg\forall x.\alpha$
- e.g. just because for every term t, we have KB $\models_c \neg\text{DirectConnect}(\text{myHome}, t)$ does not mean that KB $\models_c \forall x[\neg\text{DirectConnect}(\text{myHome}, x)]$

But may want to treat KB as providing complete information about what individuals exist

Define: KB $\models_{c2} \alpha$ iff KB $\cup$ *Negs* $\cup$ *Dc* $\cup$ *Un* $\models \alpha$

Negs is as before

Dc is <u>domain closure</u>: $\forall x[x{=}c_1 \vee ... \vee x{=}c_n]$,

Un is <u>unique names</u>: $(c_i \neq c_j)$, for $i \neq j$

where the c_i are all the constants appearing in KB (assumed finite)

Get: KB $\models_{c2} \exists x.\alpha$ iff KB $\models_{c2} \alpha[x/c]$, for some c appearing in the KB

KB $\models_{c2} \forall x.\alpha$ iff KB $\models_{c2} \alpha[x/c]$, for all c appearing in the KB

KB $\models_{c2} (c = d)$ iff c and d are the same constant

Non-monotonicity

Ordinary entailment is monotonic

If KB $\models \alpha$, then KB* $\models \alpha$, for any KB $\subseteq$ KB*

But CWA entailment is *not* monotonic

Can have KB $\models_c \alpha$, KB $\subseteq$ KB', but KB' $\not\models_c \alpha$

e.g. $\{p\} \models_c \neg q$, but $\{p, q\} \not\models_c \neg q$

Suggests study of <u>non-monotonic reasoning</u>

- start with explicit beliefs
- generate implicit beliefs non-monotonically, taking defaults into account
  - e.g. Birds fly.
- implicit beliefs may not be uniquely determined
  - vs. monotonic case: $\{\alpha \mid \text{KB} \models \alpha\}$

Will consider three approaches:

- circumscription
  - interpretations that minimize abnormality
- default logic
  - KB as facts + default rules of inference
- autoepistemic logic
  - facts that refer to what is/is not believed

Minimizing abnormality

Idea:

CWA makes the extension of all predicates as small as possible

by adding negated literals

Generalize: make extension of selected predicates as small as possible

Ab predicates used to talk about defaults

Example:

$\forall x[\text{Bird}(x) \wedge \neg\text{Ab}(x) \supset \text{Fly}(x)]$

All birds that are normal fly

Bird(chilly), ¬Fly(chilly), Bird(tweety), (chilly ≠ tweety)

Would like Fly(tweety), but KB $\not\models$ Fly(tweety)

because there is an interp ***I*** where $\Phi(\text{tweety}) \in \Phi(\text{Ab})$

Solution: consider only interps where $\Phi(\text{Ab})$ is as small as possible, relative to KB

for example: need $\Phi(\text{chilly}) \in \Phi(\text{Ab})$

Generalizes to many Ab_i predicates

Minimal Entailment

Given two interps over the same domain, I_1 and I_2

$I_1 \leq I_2$ iff $\Phi_1(\mathrm{Ab}) \subseteq \Phi_2(\mathrm{Ab})$

for every Ab predicate

$I_1 < I_2$ iff $I_1 \leq I_2$ but not $I_2 \leq I_1$

read: I_1 is more normal than I_2

Define a new version of entailment, $|=_m$, by

KB $|=_m \alpha$ iff for every I,

if I |= KB and no $I^* < I$ s.t. I^* |= KB
then I |= α.

So only requiring α to be true in interpretations satisfying KB that are <u>minimal</u> in abnormalities

Get: KB $|=_m$ Fly(tweety)

because if interp satisfies KB and is minimal, only Φ(chilly) will be in Φ(Ab).

Minimization need not produce a <u>unique</u> interpretation:

Bird(a), Bird(b), $[\neg\mathrm{Fly(a)} \vee \neg\mathrm{Fly(b)}]$

Two minimal interpretations

KB $|\neq_m$ Fly(a), KB $|\neq_m$ Fly(b), KB $|=_m$ $[\mathrm{Fly(a)} \vee \mathrm{Fly(b)}]$

Different from the CWA: no inconsistency!

Circumscription

Can achieve similar effects by leaving entailment alone, but adding a set of sentences to the KB

like CWA, but not as simple as adding $\neg$Ab(t) since we need not have constant names for abnormal individuals

Idea: say Ab, Bird, and Fly are the predicates, and suppose there are wffs $\alpha(x)$, $\beta_1(x)$, and $\beta_2(x)$ such that

KB$[\mathrm{Ab}/\alpha; \mathrm{Bird}/\beta_1; \mathrm{Fly}/\beta_2]$ is true

and $\forall x[\alpha(x) \supset \mathrm{Ab}(x)]$ is true

then want to conclude by default that

$\forall x[\alpha(x) \equiv \mathrm{Ab}(x)]$ is true.

will ensure that Ab is as small as possible

In general:

where Ab_i are the abnormality predicates
and P_i are all the other predicates,

Circ(KB) is the set of all wffs of the form

KB$[\mathrm{Ab}_1/\alpha_1; \ldots; \mathrm{Ab}_n/\alpha_n; P_1/\beta_1; \ldots ; P_m/\beta_m]$
$\wedge \forall x[\alpha_1(x) \supset \mathrm{Ab}_1(x)] \wedge \ldots \wedge \forall x[\alpha_n(x) \supset \mathrm{Ab}_n(x)]$
$\supset \forall x[\alpha_1(x) \equiv \mathrm{Ab}_1(x)] \wedge \ldots \wedge \forall x[\alpha_n(x) \equiv \mathrm{Ab}_n(x)]]$ • • •

Circumscription - 2

Theorem: If KB $\cup$ Circ(KB) $\models \alpha$ then KB $\models_m \alpha$

So this gives us a sound but incomplete method of determining minimal entailments

to get a complete version, would have to use "second order logic," which quantifies over predicates

as in: $\forall\phi[\text{KB}[\text{Ab}/\phi\ldots] \wedge \forall x(\phi(x) \supset \text{Ab}(x))$ ...

Use: guess at a "minimal" α_i and appropriate other β_i such that KB $\models$ KB[Ab/...] $\wedge\ \forall x[\alpha_i(x) \supset \text{Ab}_i(x)]$, then:

- KB[Ab/...] $\wedge \forall x[\alpha_i(x) \supset \text{Ab}_i(x)] \supset \forall x[\alpha_i(x) \equiv \text{Ab}_i(x)]$ is a member of Circ(KB)
- so KB $\cup$ Circ(KB) $\models \forall x[\alpha_i(x) \equiv \text{Ab}_i(x)]$
- since α_i was chosen to be some minimal set of abnormal individuals, it follows from KB $\cup$ Circ(KB) that these are the only instances of Ab_i
- so any other individual will have the properties of normal individuals

For the bird example, a minimal α is $(x = \text{chilly})$, for which a suitable β_1 is $\text{Bird}(x)$ and β_2 is $(x \neq \text{chilly})$.

KB $\cup$ Circ(KB) $\models \forall x[(x = \text{chilly}) \equiv \text{Ab}(x)]$

KB $\cup$ Circ(KB) $\models \neg\text{Ab}(\text{tweety})$

Fixed / variable predicates

Imagine KB as before +

$\forall x[\text{Penguin}(x) \supset \text{Bird}(x) \wedge \neg\text{Fly}(x)]$

Get: KB $\models \forall x[\text{Penguin}(x) \supset \text{Ab}(x)]$

so minimizing Ab also minimizes penguins!

Get: KB $\models_m \forall x \neg\text{Penguin}(x)$

McCarthy's definition:

Let **P** and **Q** be sets of predicates

$I_1 \leq I_2$ iff same domain and

1. $\Phi_1(P) \subseteq \Phi_2(P)$, for every $P \in$ **P** Ab predicates
2. $\Phi_1(Q) = \Phi_2(Q)$, for every $Q \notin$ **Q**

so only predicates in **Q** are allowed to vary

Get definition of $\models_m$ that is parameterized by what is minimized and what is allowed to vary

need a different definition of Circ(KB) too

In previous examples, want to minimize Ab while allowing only Fly to vary (so keep Penguin fixed)

Problems:

- need to decide what to allow to vary
- cannot conclude $\neg$Penguin(tweety) by default!

only get default $(\neg\text{Penguin}(\text{tweety}) \supset \text{Fly}(\text{tweety}))$

Default logic

Beliefs as deductive theory

explicit beliefs = axioms

implicit beliefs = theorems

least set closed under inference rules

e.g. If can prove α, $(\alpha \supset \beta)$, then infer β

Would like to generalize to default rules:

If can prove $\mathrm{Bird}(x)$, but cannot prove $\neg\mathrm{Fly}(x)$, then infer $\mathrm{Fly}(x)$.

Problem: how to characterize theorems

cannot write down a derivation as before, since we will not know when to apply default rules

no guarantee of unique set of theorems

If cannot infer p, infer q
If cannot infer q, infer p ??

Solution: default logic

no notion of theorem

instead: have extensions

sets of sentences that are "reasonable" beliefs, given facts and default rules

Extensions

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

- F is a set of sentences (facts)
- D is a set of default rules: triples $\langle \alpha, \beta, \gamma \rangle$ read as

If you can infer α and β is consistent, then infer γ

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

example: $\langle \mathrm{Bird(tweety)}, \mathrm{Fly(tweety)}, \mathrm{Fly(tweety)} \rangle$

treat $\langle \mathrm{Bird}(x), \mathrm{Fly}(x), \mathrm{Fly}(x) \rangle$ as set of rules

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

write as $\langle \alpha \Rightarrow \beta \rangle$

will see later a reason for wanting non-normal ones

A set of sentences E is an extension of $\langle F,D \rangle$ iff for every sentence π, E satisfies

$\pi \in E$ iff $F \cup \Delta \models \pi$

where $\Delta = \{\gamma \mid \langle \alpha, \beta, \gamma \rangle \in D,\ \alpha \in E, \neg\beta \notin E\}$

So, an extension E is the set of entailments of $F \cup \{\gamma\}$, where the γ are assumptions from D.

to check if E is an extension, guess at Δ and show that it satisfies the above constraint

Example

Suppose KB has

F = Bird(chilly), $\neg$Fly(chilly), Bird(tweety)
D = ‹Bird(x) $\Rightarrow$ Fly(x)›

then there is a unique extension:

Δ = Fly(tweety)

- Resulting E is an extension since tweety is the only t for this Δ such that Bird(t) $\in E$ and $\neg$Fly(t) $\notin E$.
- No other extension, since the same applies no matter what Fly(t) assumptions are in Δ.

But in general can have multiple extensions:

F = {Republican(dick), Quaker(dick)}
D = { ‹Republican(x) $\Rightarrow \neg$Pacifist(x)›, ‹Quaker(x) $\Rightarrow$ Pacifist(x)› }

conflicting defaults

Have two extensions:

E_1 has $\Delta = \neg$Pacifist(dick)
E_2 has Δ = Pacifist(dick)

Which to believe?

<u>credulous</u>: choose an extension arbitrarily

<u>skeptical</u>: believe what is common to all extensions

Can sometimes use non-normal defaults to avoid conflicts in defaults

‹Quaker(x), Pacifist(x) $\wedge \neg$Republican(x), Pacifist(x)›

but need to consider all possible interactions in defaults!

Unsupported conclusions

Definition of extension tries to eliminate facts that do not result from either F or D.

for example, we do not want Yellow(tweety) and its entailments in the extension

no unsupported conclusions

But the definition has a problem:

Suppose F = {} and D = ‹p, True, p›.

Then E = entailments of {p} is an extension

since $p \in E$ and $\neg$True $\notin E$, for above default

However, no good reason to believe p!

only support for p is default rule, which requires p itself as a prerequisite

so default rule should have no effect

Want unique extension: E = entailments of {}

Reiter's definition:

For any set S, let $\Gamma(S)$ be the least set containing F, closed under entailment, and satisfying

if ‹α, β, γ› $\in D$, $\alpha \in \Gamma(S)$, and $\neg\beta \notin S$, then $\gamma \in \Gamma(S)$.

note: not $\Gamma(S)$

A set E is an <u>extension</u> of ‹F, D› iff $E = \Gamma(E)$.

called a <u>fixed point</u> of the Γ operator

Autoepistemic logic

One disadvantage of default logic is that rules cannot be combined or reasoned about

$\langle \alpha, \beta, \gamma \rangle \overset{?}{\text{ß}} \langle \alpha, \beta, (\gamma \vee \delta) \rangle$

Solution: express defaults as <u>sentences</u> in extended language that talks about belief

for any sentence α, have another sentence $\mathbf{B}\alpha$

$\mathbf{B}\alpha$ says "I believe α": autoepistemic logic

e.g. $\forall x[\text{Bird}(x) \wedge \neg \mathbf{B} \neg \text{Fly}(x) \supset \text{Fly}(x)]$

any bird not believed to be flightless flies

These are not sentences of FOL, so what semantics and entailment?

modal logic of belief provide semantics

for here: treat $\mathbf{B}\alpha$ as if it were an new atomic wff

still get: $\forall x[\text{Bird}(x) \wedge \neg \mathbf{B} \neg \text{Fly}(x) \supset \text{Fly}(x) \vee \text{Run}(x)]$

Main property for set of implicit beliefs, E:

1. If $E \models \alpha$ then $\alpha \in E$. (entailment)
2. If $\alpha \in E$ then $\mathbf{B}\alpha \in E$. (positive introspection)
3. If $\alpha \notin E$ then $\neg \mathbf{B}\alpha \in E$. (negative introspection)

Any such set of sentences is called <u>stable</u>

Stable expansions

Given KB, possibly containing **B** operators, what is an appropriate stable set of beliefs?

want a stable set that is minimal

Moore's definition: A set of sentences E is called a <u>stable expansion</u> of KB iff it satisfies

$\pi \in E$ iff $\text{KB} \cup \Delta \models \pi$,

where $\Delta = \{\mathbf{B}\alpha \mid \alpha \in E\} \cup \{\neg \mathbf{B}\alpha \mid \alpha \notin E\}$

fixed point of another operator

analogous to the extensions of default logic

Example:

for KB = $\{\text{Bird}(\text{chilly}),\ \neg\text{Fly}(\text{chilly}),\ \text{Bird}(\text{tweety}),$
$\forall x[\text{Bird}(x) \wedge \neg \mathbf{B} \neg \text{Fly}(x) \supset \text{Fly}(x)]\}$

get a unique stable expansion containing $\text{Fly}(\text{tweety})$

As in default logic, stable expansions are not uniquely determined

KB = $\{(\neg \mathbf{B}p \supset q), (\neg \mathbf{B}q \supset p)\}$

2 stable expansions: one with p, one with q

KB = $\{(\neg \mathbf{B}p \supset p)\}$ (self-defeating default)

no stable expansions – so what to believe?

Enumerating stable expansions

Define: A wff is <u>objective</u> if it has no **B** operators

When a KB is propositional, and **B** operators only dominate objective wffs, then we can enumerate all stable expansions using the following:

1. Suppose $\mathbf{B}\alpha_1$, $\mathbf{B}\alpha_2$, ... $\mathbf{B}\alpha_n$ are all the **B** wffs in KB.
2. Replace some of these by True and the rest by ¬True in KB and simplify. Call the result KB° (it's objective).

   at most 2^n possible replacements

3. Check that for each α_i,
   - if $\mathbf{B}\alpha_i$ was replaced by True, then KB° $\models \alpha_i$
   - if $\mathbf{B}\alpha_i$ was replaced by ¬True, then KB° $\not\models \alpha_i$
4. If yes, then KB° determines a stable expansion.

   entailments of KB° are the objective part

Example:

For KB = {Bird(chilly), ¬Fly(chilly), Bird(tweety), [Bird(tweety) ∧ ¬**B**¬Fly(tweety) ⊃ Fly(tweety)], [Bird(chilly) ∧ ¬**B**¬Fly(chilly) ⊃ Fly(chilly)]}

Two **B** wffs: **B**¬Fly(tweety) and **B**¬Fly(chilly), so four replacements to try

Only one works: **B**¬Fly(tweety) → ¬True,
B¬Fly(chilly) → True

Resulting KB° has (Bird(tweety) ⊃ Fly(tweety))

More ungroundedness

Definition of stable expansion may not be strong enough

KB = {($\mathbf{B}p \supset p$)} has 2 stable expansions:

- one without p and with $\neg\mathbf{B}p$

  corresponds to KB° = {}

- one with p and $\mathbf{B}p$.

  corresponds to KB° = {p}

But why should p be believed?

only justification for having p is having $\mathbf{B}p$!

similar to problem with default logic extension

Konolige's definition:

A <u>grounded</u> stable expansion is a stable expansion that is minimal wrt to the set of sentences without **B** operators.

rules out second stable expansion

Other examples suggest that an even stronger definition is required!

can get an exact equivalence with Reiter's definition of extension in default logic