Defaults

Frames offer a simple form of default reasoning, where a slot has a certain value by inheritance unless another one is explicitly given.

Assuming that we have Dog(fido) in a KB in FOL, there are only two ways to reach the conclusion Carnivore(fido):

1. this is explicitly mentioned in KB;
2. in KB we have the universal form $\forall x.\ Dog(x) \supset Carnivore(x)$.

We are interested in expressing in FOL what we know about something in general and in particular using universals.

For example, we may say that "Bikes have two wheels" but without stating that "All bikes have two wheels" because this would rule out a bike with three wheels. A possible solution would be to say

"All bikes that are not P_1 or ... P_n have two wheels",

where P_i represent the exceptional cases. The challenge here is to characterize these cases.

We would like to make a distinction between universals that hold for all instances and generics that hold in general. Much of our knowledge about the world is generic, therefore it is important to formalize it.

Default reasoning

In general, we know that dogs are carnivores. If Fido is a dog, what are the appropriate circumstances to infer that Fido is a carnivore? We can reason as following:

Given that a P is generally a Q and knowing that P(a) is true, it is reasonable to conclude that Q(a) is true unless there is a good reason not to.

If all that we know about an individual is that it is an instance of P, then there is no reason not to conclude that it is an instance of Q.

For example, if we know that a polar bear has been playing in the mud, probably we do not want to conclude anything about its color. But if all we know is that it is a polar bear, then it is reasonable to conclude that it is white. The conclusion has no guarantee to be logically corect, it is only a reasonable default.

This form of reasoning that involves general (not universal) knowledge about a particular individual, is called <u>default reasoning</u>.

Examples of situations when we want to conclude $Q(a)$ given $P(a)$:

- General statements: Children love playing.
  Oranges are orange.
  People in a long queue become impatient.
- Lack of information to the contrary:
  No country has a president taller than 2 m.
  Children learn easily foreign languages.
- Conventions: The speed limit in a city.
  The closest shop is five minute walk (by default assume that it is open).
- Persistance: Marital status.
  The size of objects.

The list of examples is not exhaustive, but it suggests the great variety of sources of default information. Our focus is to describe exactly when it is appropriate to draw a default conclusion, in the absence of universals.

Nonmonotonicity

Ordinary deductive reasoning is monotonic, meaning that new facts produce only additional beliefs. If $KB_1 \models \alpha$ then $KB_2 \models \alpha$ for any KB_2 such that $KB_1 \subseteq KB_2$.

Default reasoning is nonmonotonic, meaning that sometimes new facts invalidate previous beliefs. For example, we believe by default that a bird flies, but if we find that the bird is an ostrich, we reconsider our belief.

I Closed-world reasoning

It is the simplest formalization of default reasoning. A finite vocabulary of predicate and constant symbols is used to represent facts about the world. But from all the valid atomic sentences, only a small fraction of them are expected to be true. The convention here is to explicitly represent the true atomic sentences and to assume that any unmentioned one is false.

Example

$$KB \begin{cases} DirectConnect(cleveland, toronto) \\ DirectConnect(toronto, chicago) \\ DirectConnect(cleveland, vancouver) \end{cases}$$

if a flight between two cities is not listed, then there is none. The closed-world assumption (CWA) is the following:

> Unless an atomic sentence is known to be true, it can be assumed to be false.

Obs. A sentence assumed to be false can be later determined to be true.

Def $KB^+ = KB \cup \{\neg p \mid p \text{ is atomic and } KB \not\models p\}$. A new form of entailment is defined as following:

$$KB \models_c \alpha \text{ iff } KB^+ \models \alpha.$$

In the previous example, KB^+ would include sentences of the form $\neg DirectConnect(c_1, c_2)$.

Consistency and completeness of knowledge

<u>Def</u>. A KB exhibits <u>consistent knowledge</u> iff there is no sentence α such that both α and $\neg\alpha$ are known.

<u>Def</u>. A KB exhibits <u>complete knowledge</u> iff for every sentence, either α or $\neg\alpha$ is known.

Knowledge can be incomplete. For example, if $KB = \{(p \vee q)\}$, neither p nor $\neg p$ can be entailed from KB.
But with the CWA, the entailment relation is complete.
For any sentence α, it holds that either $KB \models_c \alpha$ or $KB \models_c \neg\alpha$ (demonstration is by induction on the length of α).
Under CWA, whenever $KB \not\models p$ then either $KB \models \neg p$ directly or $\neg p$ is conceptually added to the KB. That means that we act as if KB represents complete knowledge.

Query evaluation

The question $KB \models_c \alpha$ reduces to questions about the literals in α:

1. $KB \models (\alpha \wedge \beta)$ iff $KB \models \alpha$ and $KB \models \beta$.
2. $KB \models \neg\neg\alpha$ iff $KB \models \alpha$.
3. $KB \models \neg(\alpha \vee \beta)$ iff $KB \models \neg\alpha$ and $KB \models \neg\beta$.
4. $KB \models_c (\alpha \vee \beta)$ iff $KB \models_c \alpha$ or $KB \models_c \beta$.
5. $KB \models_c \neg(\alpha \wedge \beta)$ iff $KB \models_c \neg\alpha$ or $KB \models_c \neg\beta$.
6. if KB^+ is consistent, then $KB \models_c \neg\alpha$ iff $KB \not\models_c \alpha$.

For example, $KB \models_c ((p \wedge q) \vee \neg(r \wedge \neg s))$ reduces to either both $KB \models_c p$ and $KB \models_c q$, or $KB \models_c \neg r$, or $KB \models_c s$.

Consistency and Generalized Closed-World Assumption (GCWA)

A consistent KB does not imply that KB^+ is also consistent.
For example, if $KB = \{(p \vee q)\}$ then KB^+ contains $\{(p \vee q), \neg p, \neg q\}$ because $KB \not\models p$ and $KB \not\models q$. So, KB^+ is not consistent.

Obs. If a KB consists of just atomic sentences (e.g. DirectConnect) or conjunctions of atomic sentences (e.g. $p \wedge q$) or disjunctions of negative literals (e.g. $\neg p \vee \neg q$), then KB^+ is consistent.

One way to preserve consistency is to restrict the application of CWA only to atoms that are "uncontroversial" (not like p and q in the example above).

Def. The generalized closed-world assumption is

$$KB^* = KB \cup \{\neg p \mid \text{for all collections of atoms } q_1, \dots, q_n, \text{ if } KB \models (p \vee q_1 \vee \dots \vee q_n) \text{ then } KB \models (q_1 \vee \dots \vee q_n)\}.$$

The entailment in GCWA is defined as following:

$$KB \models_{GC} \alpha \quad \text{iff} \quad KB^* \models \alpha.$$

Under GCWA, we will not assume that p is false if there is an entailed disjunction of atoms including p that cannot be reduced to a smaller entailed disjunction not involving p.

For example, if $KB = \{(p \vee q)\}$ then $KB \models (p \vee q)$ but $KB \not\models q$, so $\neg p \notin KB^*$ and similarly $\neg q \notin KB^*$.

If we consider an atom r, then $\neg r \in KB^*$ because $KB \models (r \vee p \vee q)$ and $KB \models (p \vee q)$.

An example of interpretation: we know that there is a direct flight from Cleveland to Dallas or Houston. As a result, we know that there is a direct flight from Cleveland to Dallas, Houston or Austin. But because there is a flight to one of the first two cities, under GCWA we will assume that there is no flight to Austin.

$KB \models (\text{DirectConnect}(\text{cleveland}, \text{dallas}) \vee \text{DirectConnect}(\text{cleveland}, \text{houston}))$

$\Rightarrow KB \models (\text{DirectConnect}(\text{cleveland}, \text{dallas}) \vee \text{DirectConnect}(\text{cleveland}, \text{houston}) \vee \text{DirectConnect}(\text{cleveland}, \text{austin}))$

$\Rightarrow \neg \text{DirectConnect}(\text{cleveland}, \text{austin}) \in KB^*$.

Entailments in GCWA are a subset of those in CWA, that is if $\neg p \in KB^*$ then $\neg p \in KB^+$.

If KB has no disjunctive knowledge, then GCWA and CWA are in complete agreement.

if KB is consistent, then KB^* is consistent.

GCWA is a weaker version of CWA that agrees with CWA in the absence of disjunctions, but remains consistent in the presence of disjunctions.

Quantifiers and Domain Closure

Let us assume that the representation language contains the predicate DirectConnect and the constants $c_1, \dots c_n$.

If KB contains only atomic sentences of the form $\text{DirectConnect}(c_i, c_j)$, then for any pair of constants c_i and c_j either $\text{DirectConnect}(c_i, c_j)$ or $\neg\text{DirectConnect}(c_i, c_j)$ is in KB^+.

Assuming that there is a city SmallTown that has no airport, then for every c_j, $\neg\text{DirectConnect}(c_j, \text{smallTown})$ is in KB^+.

If we consider the query $\neg \exists x\, DirectConnect(x, smallTown)$, under CWA neither this query nor its negation is entailed. CWA excludes any of the named cities $c_1, \dots, c_n$ flying to SmallTown, but it does not exclude other unnamed city doing so.
The easiest way to overcome this problem is to assume that the named constants are the only individuals of interest.

<u>Def</u>. The closed-world assumption with domain-closure is
$KB^{\diamond} = KB^{+} \cup \{\forall x [x = c_1 \vee \dots \vee x = c_n]\}$,
where $c_1, \dots, c_n$ are all the constant symbols appearing in KB.
The entailment in CWA with domain-closure is defined as following:

$$KB \models_{CD} \alpha \quad \text{iff} \quad KB^{\diamond} \models \alpha.$$

The main properties are:

$KB \models_{CD} \forall x \alpha$ iff $KB \models_{CD} \alpha^x_c$ for every c appearing in KB;

$KB \models_{CD} \exists x \alpha$ iff $KB \models_{CD} \alpha^x_c$ for some c appearing in KB.

Compared to CWA, in $KB^{\diamond}$ we make the additional assumption that no other objects exist apart from the named constants.
Now, under CWA with domain-closure, our query $\neg \exists x\, DirectConnect(x, smallTown)$ is entailed.

<u>Obs</u>. It is the case that $KB \models_{CD} \alpha$ or $KB \models_{CD} \neg\alpha$ for any α (with or without quantifiers).

II Circumscription

One way to express exceptional cases where a default should not apply is by using a predicate Ab (from abnormal):

$$\forall x\,[\, Bird(x) \wedge \neg Ab(x) \supset Flies(x) \,].$$

Assuming that in KB we have the additional fact:

$$Bird(chilly)$$
$$Bird(tweety)$$
$$(tweety \neq chilly)$$
$$\neg Flies(chilly)$$

we would like to conclude that Tweety flies, whereas Chilly does not.

But KB $\not\models$ Flies(tweety) because there are interpretations that satisfy KB where Flies(tweety) is false. In these interpretations, the denotation of Tweety is included in the interpretation of Ab.

The strategy for minimizing abnormality: we consider the interpretations of the KB where the interpretation of Ab is (a set) as small as possible. The default conclusions are true in models of the KB where as few of the individuals as possible are abnormal.

In the previous example, we know that Chilly is an abnormal bird, but we don't know anything about Tweety. The interpretation of Ab must include Chilly, but excludes Tweety (because nothing dictates otherwise). These technique is called circumscribing the predicate Ab.

In general, a family of predicates Ab_i is used to describe various aspects of individuals. Chilly may be in the interpretation of $Ab1$, but not in that of $Ab2$ and so on.

Minimal entailment

A new form of entailment is characterized in terms of properties of interpretations.
Let P be a fixed set of unary predicates Ab. Let $\mathfrak{I}_1 = \langle D, I_1 \rangle$ and $\mathfrak{I}_2 = \langle D, I_2 \rangle$ be interpretations over the same domain such that every constant and function is interpreted the same.
We define the relationship $\leq$ as following:

$$\mathfrak{I}_1 \leq \mathfrak{I}_2 \text{ iff for every } Pr \in P \text{ then } I_1[Pr] \subseteq I_2[Pr].$$

We say that $\mathfrak{I}_1 < \mathfrak{I}_2$ iff $\mathfrak{I}_1 \leq \mathfrak{I}_2$ and $\mathfrak{I}_2 \not\leq \mathfrak{I}_1$.

$\mathfrak{I}_1$ makes the interpretation of all Ab predicates smaller than $\mathfrak{I}_2$. In other words, $\mathfrak{I}_1$ is more normal than $\mathfrak{I}_2$.

<u>Def</u>. The minimal entailment $\models_{\leq}$ is defined as follows:
$KB \models_{\leq} \alpha$ iff for every interpretation $\mathfrak{I}$ such that $\mathfrak{I} \models KB$, either $\mathfrak{I} \models \alpha$ or there is an $\mathfrak{I}'$ such that $\mathfrak{I}' < \mathfrak{I}$ and $\mathfrak{I}' \models KB$.

In the previous example, $KB \not\models Flies(tweety)$ but $KB \models_{\leq} Flies(tweety)$.
if $\mathfrak{I} \models KB$ but $\mathfrak{I} \not\models Flies(tweety)$ then $\mathfrak{I} \models Ab(tweety)$.
We take $\mathfrak{I}'$ to be exactly $\mathfrak{I}$, except that we remove the denotation of tweety from the interpretation of Ab. Assuming that $P = \{Ab\}$, we have that $\mathfrak{I}' < \mathfrak{I}$ and $\mathfrak{I}' \models KB$.
But $\mathfrak{I}' \not\models Ab(tweety)$, so $\mathfrak{I}' \models Flies(tweety)$
Thus, in the minimal models of KB, Tweety is a normal bird
$KB \models_{\leq} \neg Ab(tweety)$ therefore $KB \models_{\leq} Flies(tweety)$.
Instead, in all the models of KB, Chilly is an abnormal bird.
In this reasoning, the only default step was to conclude that Tweety was a normal bird; the rest was ordinary deductive reasoning.

Obs. The "most normal" models of the KB may not all satisfy exactly the same sentences.

For example, suppose that the KB contains:

$$KB \begin{cases} Bird(c) \\ Bird(d) \\ \neg Flies(c) \vee \neg Flies(d) \\ \forall x\,[Bird(x) \wedge \neg Ab(x) \supset Flies(x)] \end{cases}$$

In any model of the KB, the interpretation of Ab must contain either the denotation of c or the denotation of d. Any model containing other abnormal individuals would not be minimal (including both c and d).

So, in any minimal model $\mathfrak{I} \models KB$, we have either $\mathfrak{I} \models Ab(c)$ or $\mathfrak{I} \models Ab(d)$.

if $\mathfrak{I} \models Ab(c)$ then $KB \not\models_{\leq} Flies(c)$. Similarly, if $\mathfrak{I} \models Ab(d)$ then $KB \not\models_{\leq} Flies(d)$.

We cannot conclude by default that c is a normal bird, nor that d is. But we can conclude by default that one of them is: $KB \models_{\leq} Flies(c) \vee Flies(d)$.

Obs. CWA and GCWA have a different behavior.

Because neither $KB \not\models Ab(c)$ nor $KB \not\models Ab(d)$, it results that

$$KB^+ \supset KB \cup \{\neg Ab(c), \neg Ab(d)\}.$$

So, $KB \models_c (Flies(c) \wedge Flies(d))$, that is KB^+ is not consistent.

On the other hand, under GCWA $\neg Ab(c) \notin KB^*$ and $\neg Ab(d) \notin KB^*$.

$[KB \models Ab(c) \vee \neg Bird(c) \vee Flies(c)$ but $KB \not\models \neg Bird(c) \vee Flies(c)]$

So, under GCWA we cannot conclude anything about Flies(c) or Flies(d) (or their disjunction).

In the circumscription case, one model of the KB is prefered to another one if it has less abnormal individuals. Assuming that we have the statements: Richard Nixon was both quaker (thus implicitly pacifist) and republican (thus implicitly not pacifist), we have the following KB:

$$KB \left[\begin{array}{l} Republican(nixon) \wedge Quaker(nixon) \\ \forall x\, [\, Republican(x) \wedge \neg Ab_2(x) \supset \neg Pacifist(x) \,] \\ \forall x\, [\, Quaker(x) \wedge \neg Ab_3(x) \supset Pacifist(x) \,] \end{array} \right.$$

if we circumscribe the predicates Ab2 and Ab3, we have two minimal models

$$J_1 \models Ab_2(nixon) \quad \text{and} \quad J_1 \models Pacifist(nixon)$$

$$J_2 \models Ab_3(nixon) \quad \text{and} \quad J_2 \models \neg Pacifist(nixon)$$

So, $KB \not\models_{\leq} Pacifist(nixon)$ and $KB \not\models_{\leq} \neg Pacifist(nixon)$.

if, for example, we give priority to religious convictions rather than to political ones, we can express it by prioritized circumscription, where we prefer the (minimal) model that minimizes Ab3.

III Default logic

It provides a mechanism that explicitly specifies which sentences should be added to the KB, maintaining consistency. For example, if Bird(t) is entailed by the KB, we might want to add the default assumption Flies(t), if it is consistent to do so.

In default logic, a KB consists of two parts: a set $\mathcal{F}$ of first-order sentences and a set $\mathcal{D}$ of default rules, which are specifications of what assumptions can be made and when.

The role of the default logic is to specify the following:
- the appropriate set of implicit beliefs that incorporate the facts in $\mathcal{F}$;
- as many default assumptions as possible, given the default rules in $\mathcal{D}$;
- the logical entailments inferred from the implicit beliefs and the default assumptions.

Default rules

A default rule is written in the form $\langle \alpha : \beta / \delta \rangle$, where α is the prerequisite, β is the justification and δ is the conclusion. δ is considered to be true if α is true and it is consistent to believe β (that is $\neg\beta$ is not true).

$$\langle Bird(tweety) : Flies(tweety) / Flies(tweety) \rangle$$

A rule where the justification and conclusion are the same is called a normal default rule and it is written as $Bird(tweety) \Rightarrow Flies(tweety)$.

A rule can be formulated using free variables:

$\langle Bird(x) : Flies(x) / Flies(x) \rangle$ represents the set of all its instances, formed by replacing x by a ground term.

Default extensions

Given a default theory $KB = (\mathcal{F}, \mathcal{D})$, what are the sentences that should be believed?

We define an extension of the theory as a reasonable set of beliefs given a default theory.

Def. A set of sentences $\mathcal{E}$ is an extension of a default theory $(\mathcal{F}, \mathcal{D})$ if for every sentence π we have:

$$\pi \in \mathcal{E} \text{ iff } \mathcal{F} \cup \{\delta \mid \langle \alpha : \beta / \delta \rangle \in \mathcal{D}, \alpha \in \mathcal{E}, \neg\beta \notin \mathcal{E}\} \models \pi .$$

Thus, an extension is the set of all entailments of $\mathcal{F} \cup \Delta$, where Δ is a suitable set of assumptions.

<u>Obs</u>. The definition of $\mathcal{E}$ does not say how to find an $\mathcal{E}$, but $\mathcal{E}$ is completely characterized by its set of applicable assumptions Δ.

<u>Example</u>

$$\mathcal{F} = \{ \text{Bird}(\text{tweety}), \text{Bird}(\text{chilly}), \neg \text{Flies}(\text{chilly}) \}$$
$$D = \{ \text{Bird}(x) \Rightarrow \text{Flies}(x) \}$$

Let $\mathcal{E} = \mathcal{F} \cup \{ \text{Flies}(\text{tweety}) \}$.

Flies(tweety) is the only assumption applicable to $\mathcal{E}$.

$$\left.\begin{array}{l} \text{Bird}(\text{tweety}) \in \mathcal{E} \\ \neg \text{Flies}(\text{tweety}) \notin \mathcal{E} \end{array}\right\} \Rightarrow \text{Flies}(\text{tweety}) \text{ is applicable.}$$

Flies(t) is not applicable for any other t (in our example t could be only chilly).

Thus, Flies(tweety) is the only applicable assumption, so $\mathcal{E}$ is an extension (it can be proven that it is the only one).

<u>Obs</u>. An extension $\mathcal{E}$ of a default theory $(\mathcal{F}, D)$ is inconsistent iff $\mathcal{F}$ is inconsistent.

<u>Multiple extensions</u>

Consider the following default theory:

$$\mathcal{F} = \{ \text{Republican}(\text{nixon}), \text{Quaker}(\text{nixon}) \}$$
$$D = \{ \text{Republican}(x) \Rightarrow \neg \text{Pacifist}(x), \text{Quaker}(x) \Rightarrow \text{Pacifist}(x) \}.$$

Let $\mathcal{E}_1$ be the extension characterized by the assumption Pacifist(nixon) and $\mathcal{E}_2$ characterized by the assumption $\neg$Pacifist(nixon).

$\mathcal{E}_1$ and $\mathcal{E}_2$ are extensions because their assumptions are applicable and there are no other applicable ones (for $t \neq$ nixon).

The empty set of assumptions does not give an extension, because both Pacifist(nixon) and $\neg$Pacifist(nixon) would be applicable. For any other extensions, assumptions would be of the form Pacifist(t) or $\neg$Pacifist(t), but none are applicable for $t \neq$ nixon.

Thus, $\mathcal{E}_1$ and $\mathcal{E}_2$ are the only extensions possible.

On the basis of what we know, either Nixon is a pacifist or he is not a pacifist are reasonable beliefs. There are two options:

1. a skeptical reasoner will only believe those sentences that are common to all extensions of the default theory;
2. a credulous reasoner will simply choose an extension as a set of sentences to believe.

In some cases, the existance of multiple extensions is an indication that we have not said enough to make a reasonable decision.

In the previous example, we may want to say that the default about Quakers should apply only to individuals that are not politically active. If we add in $\mathcal{F}$ the fact

$$\forall x\,[\,\text{Republican}(x) \supset \text{Political}(x)\,],$$

we can replace the rule Quaker(x) $\Rightarrow$ Pacifist(x) by a non-normal one:

$$\frac{\text{Quaker}(x) : [\,\text{Pacifist}(x) \wedge \neg\text{Political}(x)\,]}{\text{Pacifist}(x)}.$$

For ordinary Quakers, the assumption is that they are pacifists. But for Quaker Republicans like Nixon, we assume that they are not pacifists.

If we replace $\forall x\,[\,\text{Republican}(x) \supset \text{Political}(x)\,]$ by the default rule Republican(x) $\Rightarrow$ Political(x), then we have two extensions:

- one characterized by the assumptions {$\neg$Pacifist(nixon), Political(nixon)};
- one characterized by the assumption {Pacifist(nixon)}.

Resolving conflicts among default rules is crucial when we deal with concept hierarchies.

For example, for the following KB:

$$\mathcal{F} = \{ \forall x [\, Penguin(x) \supset Bird(x) \,],\ Penguin(chilly) \}$$

$$\mathcal{D} = \{ Bird(x) \Rightarrow Flies(x),\ Penguin(x) \Rightarrow \neg Flies(x) \}$$

we have two extensions: one where Chilly is assumed to fly and one where Chilly is assumed not to fly.

The default that penguins do not fly should preempt the more general default that birds fly.

$$\frac{Bird(x) : [\, Flies(x) \wedge \neg Penguin(x) \,]}{Flies(x)}$$

Unlike defaults in an inheritance mechanism, the default logic do not automatically prefer the most specific defaults.