G53KRR handout on defaults.

Defaults or default rules as opposed to normal or categorical rules are ways of drawing conclusions which are justified unless there is some explicit reason to believe otherwise. So normal rule will say 'if x is a natural number then it is greater or equal to 0' and this is really true without exceptions for all natural numbers. A default rule would say 'if x is a bird then it can fly (unless there are good reasons to believe otherwise)'. In other words, if all we know about x is that it is a bird, then it is reasonable to conclude that it can fly. Later however we may discover that it is a special kind of bird which does not fly.

Non-monotonicity In classical reasoning, entailment is *monotonic*: if $KB_1 \models \phi$ and $KB_1 \subseteq KB_2$, then $KB_2 \models \phi$. In other words, if ϕ is entailed by KB_1 and we add more sentences to KB_1, ϕ will still be entailed by the resulting knowledge base; the larger the knowledge base, the more consequences it has: if $KB_1 \subseteq KB_2$ then $Consequences(KB_1) \subseteq Consequences(KB_2)$.

Default reasoning is *nonmonotonic*. If we have $KB_1 = \{$ 'Birds normally can fly', 'Tweety is a bird' $\}$ then we can derive by default that Tweety can fly. However, if we learn more about Tweety, for example that it is a penguin, then 'Tweety can fly' no longer follows even by default.

The question is, how to make this work formally (define what are valid default consequences)? In these two lectures, consider three approaches: closed-world assumption, circumscription, default logic.

Closed-world reasoning *Closed-world assumption* (CWA): if an atomic sentence is not in the knowledge base, it is assumed to be false. (Like negation as failure in production rule systems: if a fact is not in the working memory, then its negation matches/is assumed to be true.) The corresponding entailment $\models_{CWA}$:

$$KB \models_{CWA} \phi \;\Leftrightarrow\; KB^+ \models \phi$$

where $KB^+ = KB \cup \{\neg p : p \text{ is atomic and } KB \not\models p\}$.

If $KB = \{Bird(t)\}$ then $\neg Penguin(t)$ follows under CWA.

Problems: if $KB = \{p \vee q\}$, KB^+ is inconsistent since it contains $p \vee q$, $\neg p$ and $\neg q$, so everything follows from it.

Generalised CWA is a fix for this:

$$KB \models_{GCWA} \phi \;\Leftrightarrow\; KB^* \models \phi$$

where $KB^* = KB \cup \{\neg p : p$ is atomic and for all collections of atoms $q_1, \ldots, q_n$, if $KB \models p \vee q_1 \vee \ldots \vee q_n$, then $KB \models q_1 \vee \ldots \vee q_n\}$.

CWA with domain closure: only explicitly named individuals are assumed to exist:

$$KB \models_{CD} \phi \;\Leftrightarrow\; KB^\diamond \models \phi$$

where $KB^\diamond = KB^+ \cup \forall x(x = c_1 \vee \ldots \vee x = c_n)$ where $c_1, \ldots, c_n$ are all the constant symbols appearing in KB.

Under CWA with domain closure, if there is no fact $P(a)$ in the knowledge base then it entails by default $\neg \exists x P(x)$.

Unique name assumption: $c \neq c'$ for any two distinct constants c, c'.

Reasoning under CWA is constructive and reasonably efficient.

Circumscription (John McCarthy). This is a generalisation of CWA: for default entailment, consider not all models of KB but only those where the set of exceptions is made as small as possible. Namely, consider a predicate *Ab* (for abnormal) and the formulation of a default rule as

$$\forall x(Bird(x) \wedge \neg Ab(x) \supset Flies(x))$$

and say that a conclusion follows by default if it is entailed on all interpretations where the extension of *Ab* is as small as possible. (This is called *circumscribing Ab* and the approach is called *circumscription*.) We need one *Ab* for every default rule, because a bird which is abnormal with respect to flying may be normal with respect to having two legs etc.

Let A be the set of *Ab* predicates we want to minimise. Let $M_1 = (D, I_1)$ and $M_2 = (D, I_2)$ be two interpretations over the same domain such that every constant and function are interpreted the same way.

$$M_1 \leq M_2 \;\Leftrightarrow\; \forall Ab \in A\ (I_1(Ab) \subseteq I_2(Ab))$$

$M_1 < M_2$ if $M_1 \leq M_2$ but not $M_2 \leq M_1$. (There are strictly fewer abnormal things in M_1).

Minimal entailment: $KB \models_{\leq} \phi$ iff for all interpretations M which make KB true, either $M \models \phi$ or M is not minimal (exists M' such that $M' < M$ and $M' \models KB$).

Example:

$$KB = \{Bird(chilly), Bird(tweety), (tweety \neq chilly), \neg Flies(chilly), \forall x(Bird(x) \land \neg Ab(x) \supset Flies(x))\}$$

$KB \not\models Flies(tweety)$ but $KB \models_{\leq} Flies(tweety)$. This knowledge base has a unique minimal extension for Ab, $I(Ab) = \{chilly\}$. This is not always the case. For example, if instead of $\neg Flies(chilly)$ it had $\neg Flies(chilly) \lor \neg Flies(tweety)$ there would be two minimal extensions of *Ab*: one where Tweety is abnormal and another where Chilly is abnormal.

Different from CWA: cannot replicate this effect by adding a fixed set of negated atomic sentences to KB.

Circumscription has constructive reasoning procedures but complexity is high.

Default logic Roy Reiter. A *default rule* consists of a *prerequisite* α, *justification* β, *conclusion* γ and says 'if α holds and it is consistent to believe β, then believe γ': $\dfrac{\alpha : \beta}{\gamma}$

For example:

$$\frac{Bird(x) : Flies(x)}{Flies(x)}$$

Default rules where justification and conclusion are the same are called *normal default rules* and are writted $Bird(x) \Rightarrow Flies(x)$.

Given a *default theory* $KB = \{F, D\}$, where F is a finite set of first order sentences and D is a finite set of default rules, what is the set of reasonable beliefs (extension of the default theory)?

E is an *extension* of (F, D) iff for every sentence π,

$$\pi \in E \;\Leftrightarrow\; F \cup \{\gamma \mid \frac{\alpha : \beta}{\gamma} \in D, \alpha \in E, \neg\beta \notin E\} \models \pi$$

where $\{\gamma \mid \frac{\alpha:\beta}{\gamma} \in D, \alpha \in E, \neg\beta \notin E\}$ is a set of *applicable assumptions* for this extension.

Example: $F = \{Bird(tweety), Bird(chilly), \neg Flies(chilly)\}$, $D = \{Bird(x) \Rightarrow Flies(x)\}$. The only applicable assumption is $Flies(tweety)$ (prerequisite in F hence in E, negation of justification not in E). For Chilly, negation of justification is in E, and for no other object o we can have $Bird(o)$ in E. So $E = F \cup \{Flies(tweety)\}$ so the theory entails $Flies(tweety)$.

Extensions are not defined constructively and they are not unique: for example,

Facts: $F = \{Republican(dick), Quaker(dick)\}$

Default rules: $Republican(x) \Rightarrow \neg Pacifist(x)$, $Quaker(x) \Rightarrow Pacifist(x)$.

Two extensions: E_1 where $Pacifist(dick)$, E_2 where $\neg Pacifist(dick)$. Can be forced to make only one extension using a non-normal default rule:

$$\frac{Quaker(x) : Pacifist(x) \land \neg MemberOfPoliticalParty(x)}{Pacifist(x)}$$

and a rule $\forall x(Republican(x) \supset MemberOfPoliticalParty(x))$.

skeptical reasoner will only believe sentences which belong to all extensions of the default theory; *credulous* reasoner will choose an arbitrary extension.