Introduction to knowledge representation and reasoning

We now look briefly at how knowledge about the world might be represented and reasoned with.

Aims:

- To introduce *semantic networks* and *frames* for knowledge representation.
- To see how *inheritance* can be applied as a reasoning method.
- To look at the use of *rules* for knowledge representation, along with *forward chaining* and *backward chaining* for reasoning.

Reading: *The Essence of Artificial Intelligence*, Alison Cawsey. Prentice Hall, 1998.

Knowledge representation

The “manipulation of knowledge” seems to be at the heart of what we as intelligent beings do.

To try to model this process in an agent we:

- *represent* knowledge using *symbol structures*, and;
- perform *formalised* versions of reasoning.

This means that we need carefully specified *languages* for the representation of knowledge.

Requirements for a knowledge representation language

First, we need **representational adequacy**.

Can I represent the pieces of knowledge I need to?

Propositional logic might well fail this test, although *predicate logic* seems better and is indeed a standard tool.

Requirements for a knowledge representation language

Or more subtly:

Can I represent the pieces of knowledge I need to in such a way that reasoning can be automated?

English is excellent and highly expressive in representing knowledge:

“Ophelia believes that all sensible people dislike eating pies”

However automating reasoning based on English language representations is just about impossible at present.

How would we write a program that takes this statement and when told *“Neddy is really jolly sensible”* and *“Neddy is a funny sort of person”* infers that *“Ophelia believes Neddy dislikes eating pies”*?

Requirements for a knowledge representation language

On the other hand:

$$\texttt{person}(\texttt{neddy})$$

$$\texttt{sensible}(\texttt{neddy})$$

$$\forall x\ \texttt{sensible}(x) \wedge \texttt{person}(x) \rightarrow (\forall y\ \texttt{pie}(y) \rightarrow \texttt{dislikes}(x, y))$$

is something for which reasoning can be automated.

Syntax and semantics

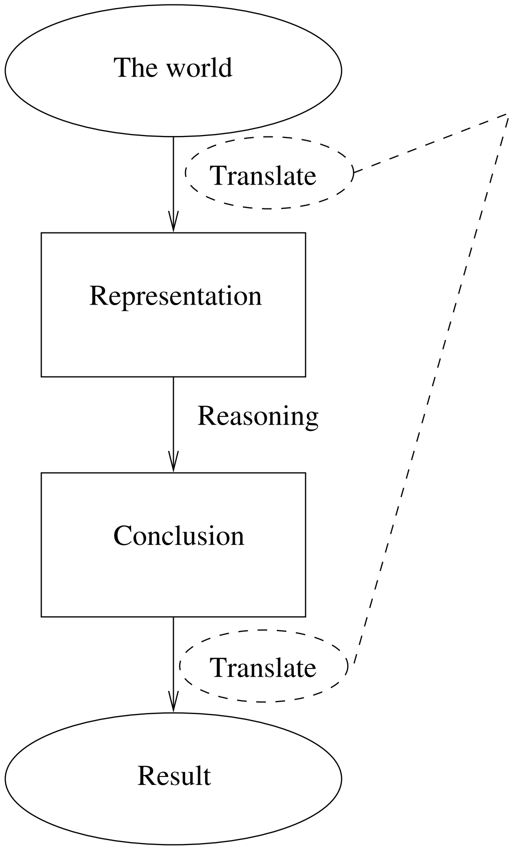
In addition to needing an expressive language, the language needs to be clearly defined:

- **Syntax:** defining when a statement in the language is well-formed.
- **Semantics**: specifying what a statement in the language *means*.

English is again not good here from the point of view of automation. Logic is again preferable.

If possible, we also want the representation to be *natural* in the sense that it is reasonably easy to understand and deal with.

Syntax and semantics



To translate confidently we need well-defined syntax and semantics

Does

`dislikes`(`neddy`, `pie`)

mean Neddy dislikes all pies, or that he's just taken a dislike to an individual pie?

Inferential adequacy and inferential efficiency

We also need to know that we can infer the things of interest:

- It is not always possible, and it's certainly not desirable, to store all knowledge as explicit facts.

  Knowing that *"all dogs smell bad"* should allow us to infer that *"fido smells bad" etc*. We don't want to store a piece of knowledge for every possible dog.
- However, more complex inferences are likely to take longer.

So as usual, there is a trade-off.

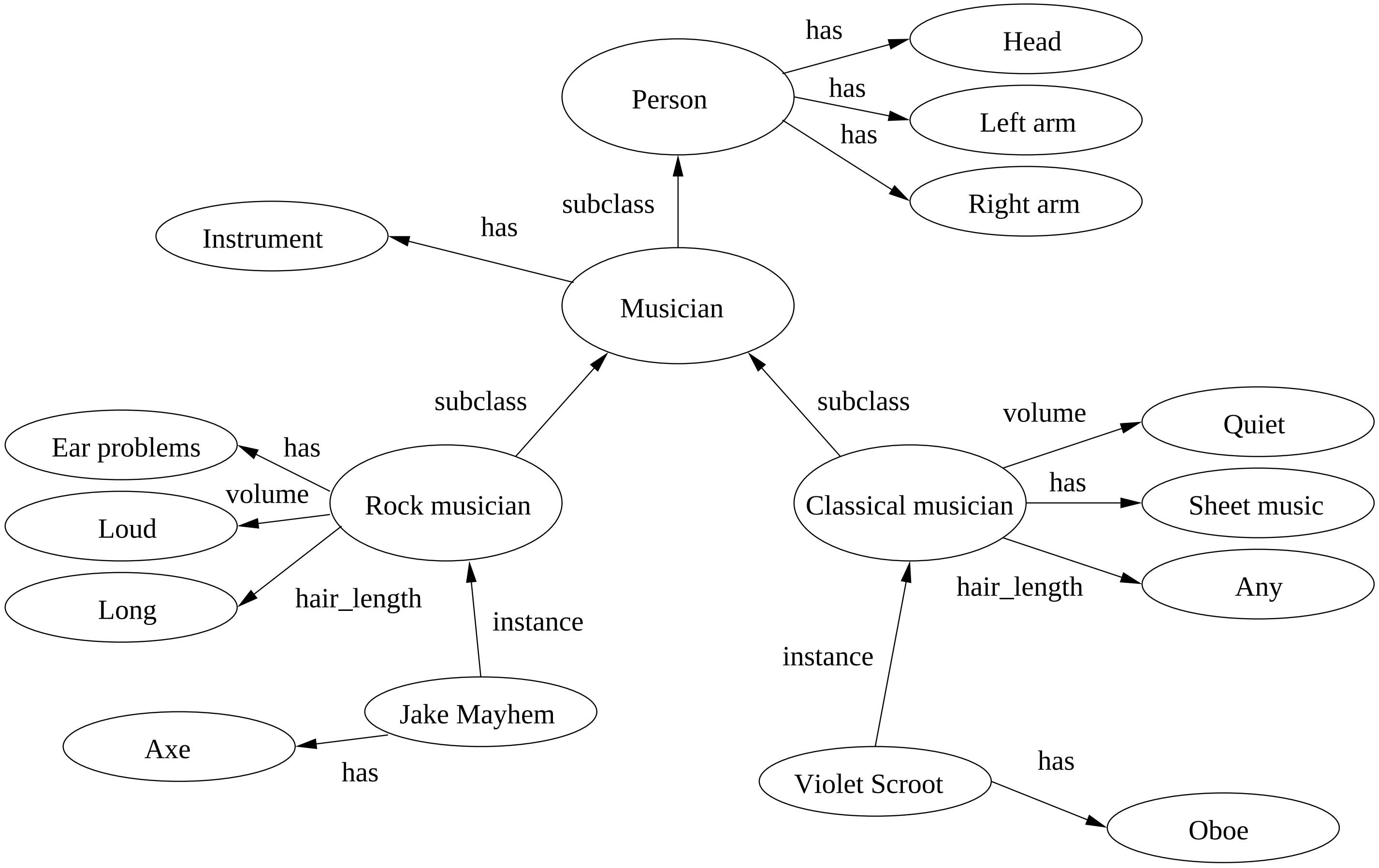
Frames and semantic networks

Frames and semantic networks represent knowledge in the form of *classes of objects* and *relationships between them*:

- the *subclass* and *instance* relationships are emphasised;
- we form *class hierarchies* in which *inheritance* is supported and provides the main *inference* mechanism;
- as a result inference is quite limited;
- we need to be extremely careful about *semantics*.

The only major difference between the two ideas is *notational*.

Example of a semantic network



Frames

Frames once again support inheritance through the subclass relationship.

Rock musician

subclass: Musician
has: ear problems
hairlength: long
volume: loud

Musician

subclass: Person
has: instrument

has, hairlength, volume *etc* are “slots”.

long, loud, instrument *etc* are “slot values”.

These are a direct predecessor of object-oriented programming languages.

Defaults

Both approaches to knowledge representation are able to incorporate *defaults*:

Rock musician

subclass: Musician
has: ear problems
* hairlength: long
* volume: loud

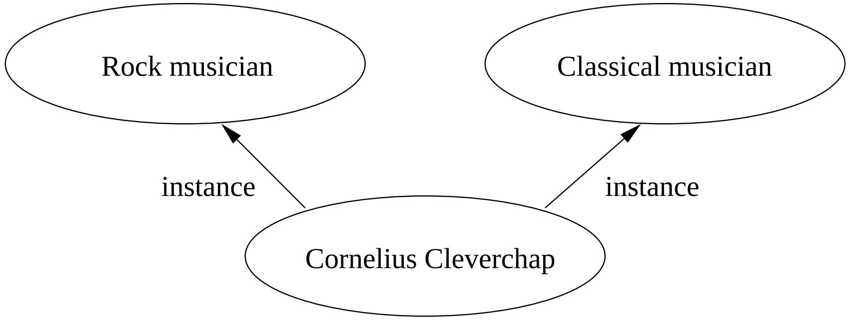
Dementia Evilperson

subclass: Rock musician
hairlength: short
image: gothic

Starred slots are *typical* values associated with subclasses and instances, but can be overridden.

Multiple inheritance

Both approaches can incorporate *multiple inheritance*, at a cost:



- what is `hairlength` for `Cornelius` if we're trying to use inheritance to establish it?
- this can be overcome initially by specifying which class is inherited from in preference when there's a conflict;
- but the problem is still not entirely solved—what if we want to prefer inheritance of some things from one class, but inheritance of others from a different one?

Other issues

- Slots and slot values can themselves be frames. For example `Dementia` may have an instrument slot with the value `Electric harp`, which itself may have properties described in a frame.
- Slots can have *specified attributes*. For example, we might specify that `instrument` can have multiple values, that each value can only be an instance of `Instrument` that each value has a slot called `owned_by` and so on.
- Slots may contain arbitrary pieces of program. This is known as *procedural attachment*. The fragment might be executed to return the slot's value, or update the values in other slots *etc*.

Rule-based systems

A rule-based system requires three things:

1. A set of `if-then` rules. These denote specific pieces of knowledge about the world.

   They should be interpreted similarly to logical implication, rather than the programming construct. In particular a collection of such rules doesn't necessarily imply a sequence.

   Such rules denote *what to do* or *what can be inferred* under given circumstances.

2. A collections of *facts* denoting what the system regards as currently true about the world.

3. An interpreter able to apply the current rules in the light of the current facts.

Forward chaining

The first of two basic kinds of interpreter begins with established facts and then applies rules to them.

This is a *data-driven* process. It is appropriate if we know the *initial facts* but not the required conclusion.

Example: XCON—used for configuring VAX computers.

In addition:

- we maintain a *working memory*, typically of what has been inferred so far;
- rules are often *condition-action rules*, where the right-hand side specifies an action such as adding or removing something from working memory, printing a message *etc*;
- in some cases actions might be entire program fragments.

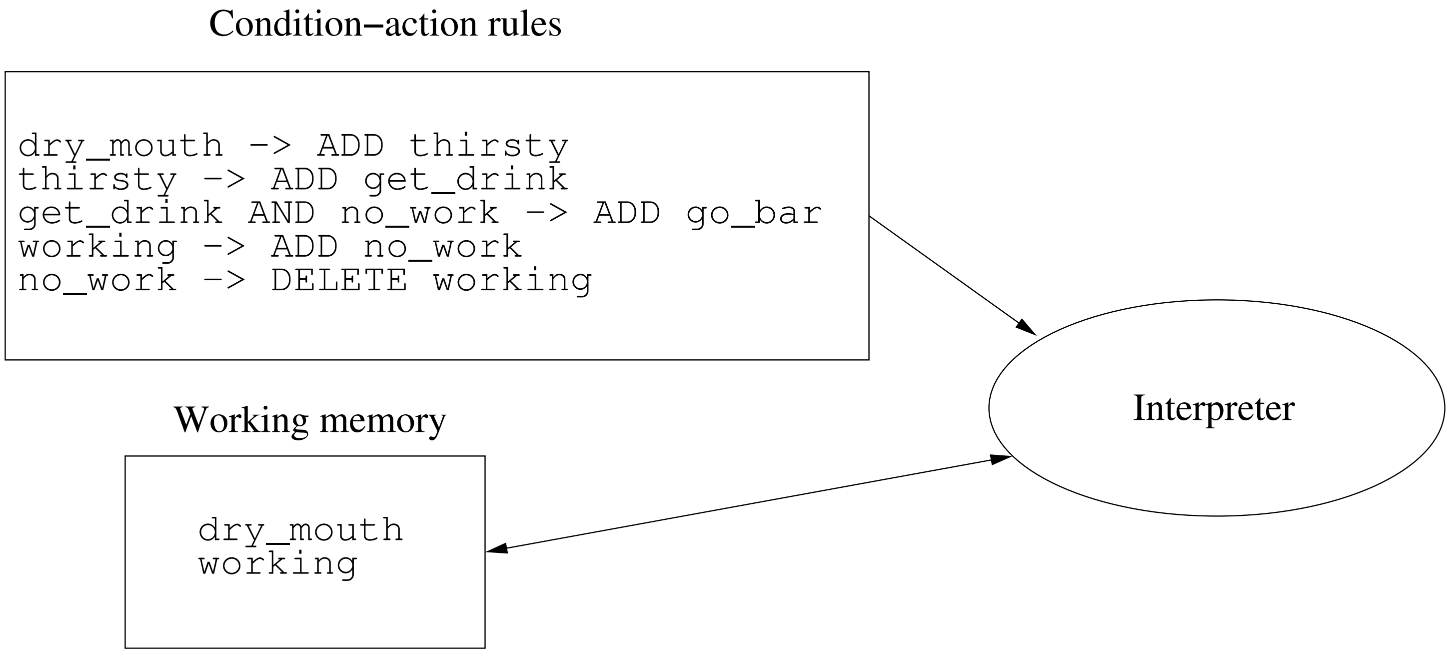
Forward chaining

The basic algorithm is:

1. find all the rules that can fire, based on the current working memory;
2. select a rule to fire. This requires a *conflict resolution strategy*;
3. carry out the action specified, possibly updating the working memory.

Repeat this process until either no rules can be used or a “halt” appears in the working memory.

Example



Condition–action rules

```
dry_mouth -> ADD thirsty
thirsty -> ADD get_drink
get_drink AND no_work -> ADD go_bar
working -> ADD no_work
no_work -> DELETE working
```

Working memory

```
dry_mouth
working
```

Interpreter

Example

Progress is as follows:

1. The rule

   `dry_mouth` → ADD `thirsty`

   fires adding `thirsty` to working memory.
2. The rule

   `thirsty` → ADD `get_drink`

   fires adding `get_drink` to working memory.
3. The rule

   `working` → ADD `no_work`

   fires adding `no_work` to working memory.
4. The rule

   `get_drink` AND `no_work` → ADD `go_bar`

   fires, and we establish that it’s time to go to the bar.

Conflict resolution

Clearly, in any more realistic system we expect to have to deal with a scenario where two or more rules can be fired at any one time:

- which rule we choose can clearly affect the outcome;
- we might also want to attempt to avoid inferring an abundance of useless information.

We therefore need a means of resolving such conflicts.

Conflict resolution

Common *conflict resolution* strategies are:

- prefer rules involving more recently added facts;
- prefer rules that are *more specific*. For example

  `patient_coughing` → ADD `lung_problem`

  is more general than

  `patient_coughing` AND `patient_smoker` → ADD `lung_cancer`.

  This allows us to define exceptions to general rules;
- allow the designer of the rules to specify priorities;
- fire all rules simultaneously—this essentially involves following all chains of inference at once.

Reason maintenance

Some systems will allow information to be removed from the working memory if it is no longer *justified*.

For example, we might find that

`patient_coughing`

and

`patient_smoker`

are in working memory, and hence fire

`patient_coughing` AND `patient_smoker` → ADD `lung_cancer`

but later infer something that causes `patient_coughing` to be withdrawn from working memory.

The justification for `lung_cancer` has been removed, and so it should perhaps be removed also.

Pattern matching

In general rules may be expressed in a slightly more flexible form involving *variables* which can work in conjunction with *pattern matching*.

For example the rule

$$\texttt{coughs}(X) \text{ AND } \texttt{smoker}(X) \rightarrow \text{ADD } \texttt{lung\_cancer}(X)$$

contains the variable X.

If the working memory contains $\texttt{coughs}(\texttt{neddy})$ and $\texttt{smoker}(\texttt{neddy})$ then

$$X = \texttt{neddy}$$

provides a match and

$$\texttt{lung\_cancer}(\texttt{neddy})$$

is added to the working memory.

Backward chaining

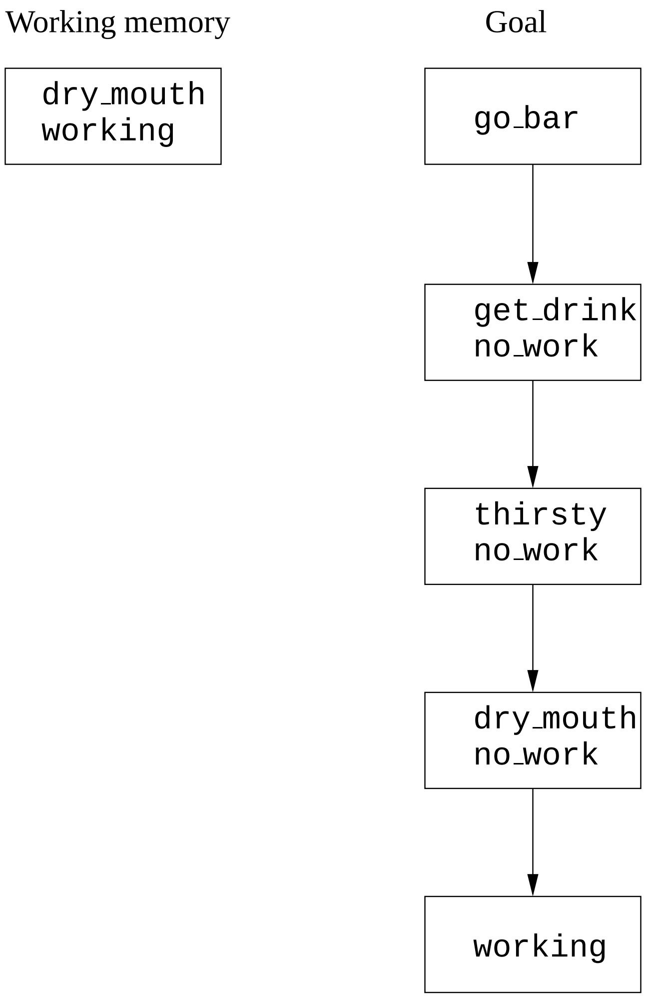
The second basic kind of interpreter begins with a *goal* and finds a rule that would achieve it.

It then works backwards, trying to achieve the resulting earlier goals in the succession of inferences.

Example: MYCIN—medical diagnosis with a small number of conditions.

This is a *goal-driven* process. If you want to *test a hypothesis* or you have some idea of a likely conclusion it can be more efficient than forward chaining.

Example



Working memory

`dry_mouth`
`working`

Goal

`go_bar`

`get_drink`
`no_work`

To establish `go_bar` we have to establish `get_drink` and `no_work`. These are the new goals.

`thirsty`
`no_work`

Try first to establish `get_drink`. This can be done by establishing `thirsty`.

`dry_mouth`
`no_work`

`thirsty` can be established by establishing `dry_mouth`. This is in the working memory so we're done.

`working`

Finally, we can establish `no_work` by establishing `working`. This is in the working memory so the process has finished.

Example with backtracking

If at some point more than one rule has the required conclusion then we can *backtrack*.

Example: *Prolog* backtracks, and incorporates pattern matching. It orders attempts according to the order in which rules appear in the program.

Example: having added

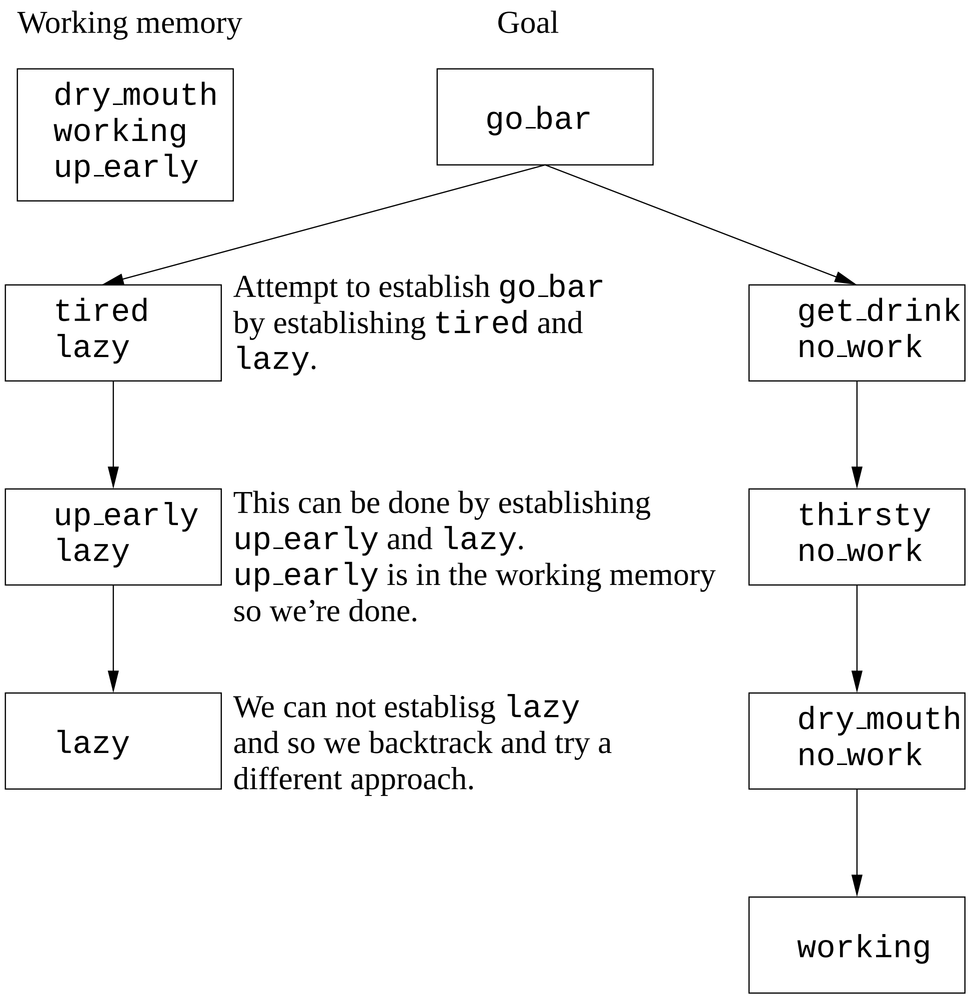
`up_early` → ADD `tired`

and

`tired` AND `lazy` → ADD `go_bar`

to the rules, and `up_early` to the working memory:

Example with backtracking



Working memory

`dry_mouth`
`working`
`up_early`

Goal

`go_bar`

`tired`
`lazy`

Attempt to establish `go_bar` by establishing `tired` and `lazy`.

`up_early`
`lazy`

This can be done by establishing `up_early` and `lazy`. `up_early` is in the working memory so we're done.

`lazy`

We can not establisg `lazy` and so we backtrack and try a different approach.

`get_drink`
`no_work`

`thirsty`
`no_work`

`dry_mouth`
`no_work`

`working`

Process proceeds as before