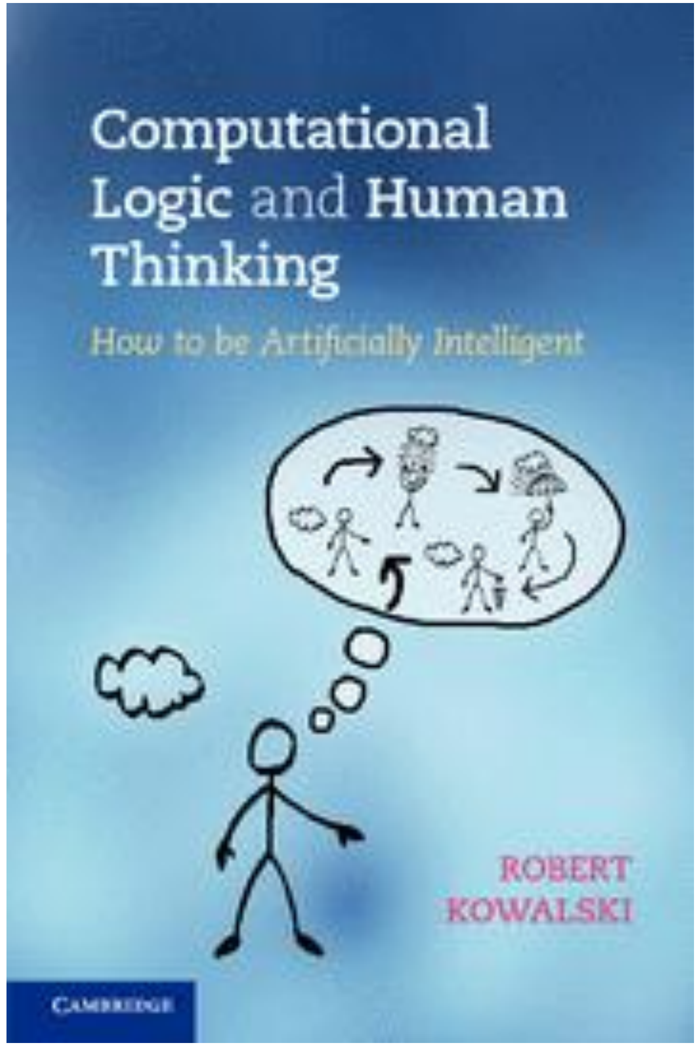
Artificial Intelligence and Human Thinking

Robert Kowalski
Imperial College London


Computational Logic and Human Thinking
How to be Artificially Intelligent
ROBERT KOWALSKI
CAMBRIDGE

AI tools and techniques

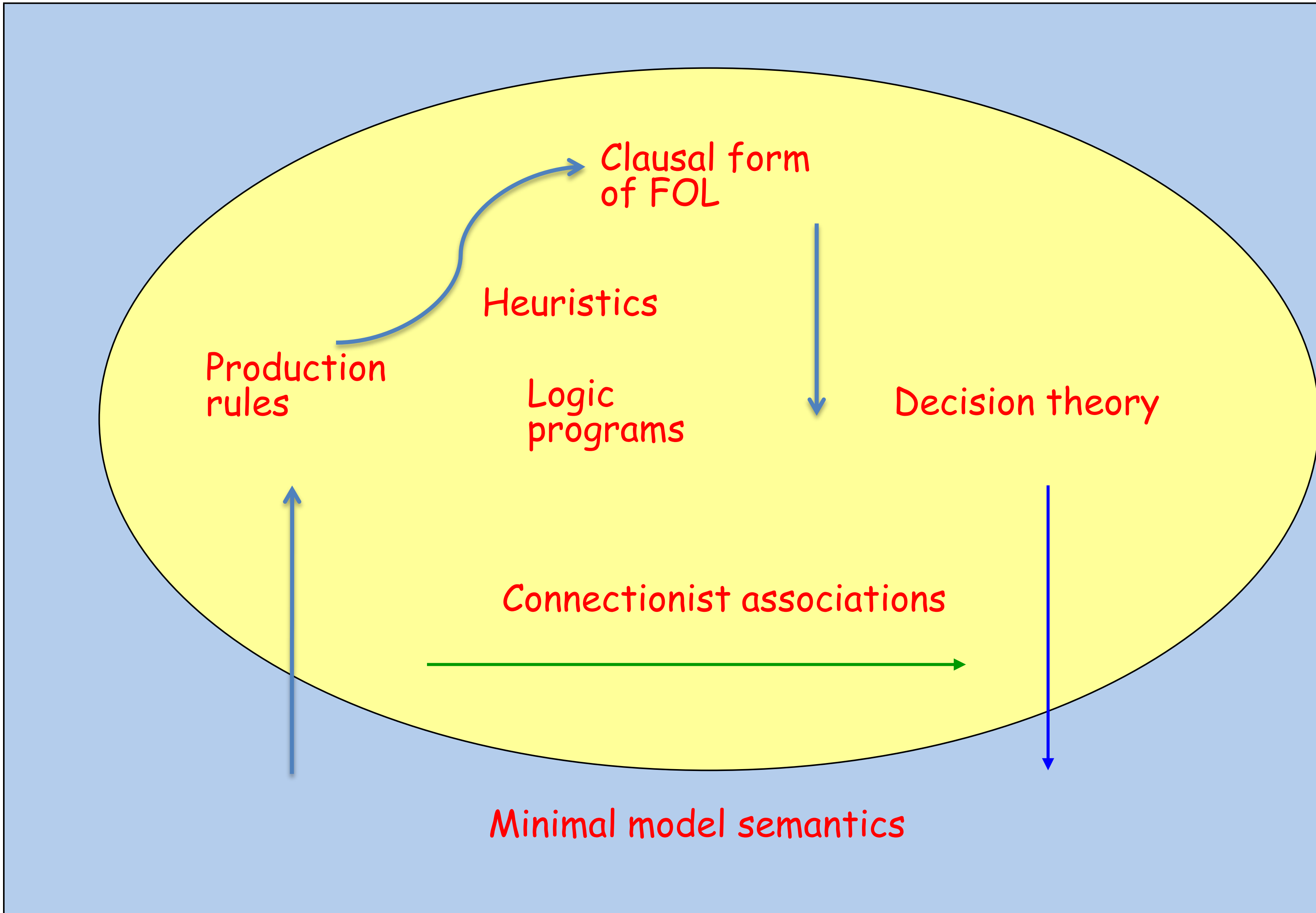
- can be reconciled and combined
  - Logic
  - Procedural representations
  - Heuristics
  - Production Systems
  - Bayesian networks
  - Connectionism

- can help people
  - make better decisions
  - communicate more effectively with other people.

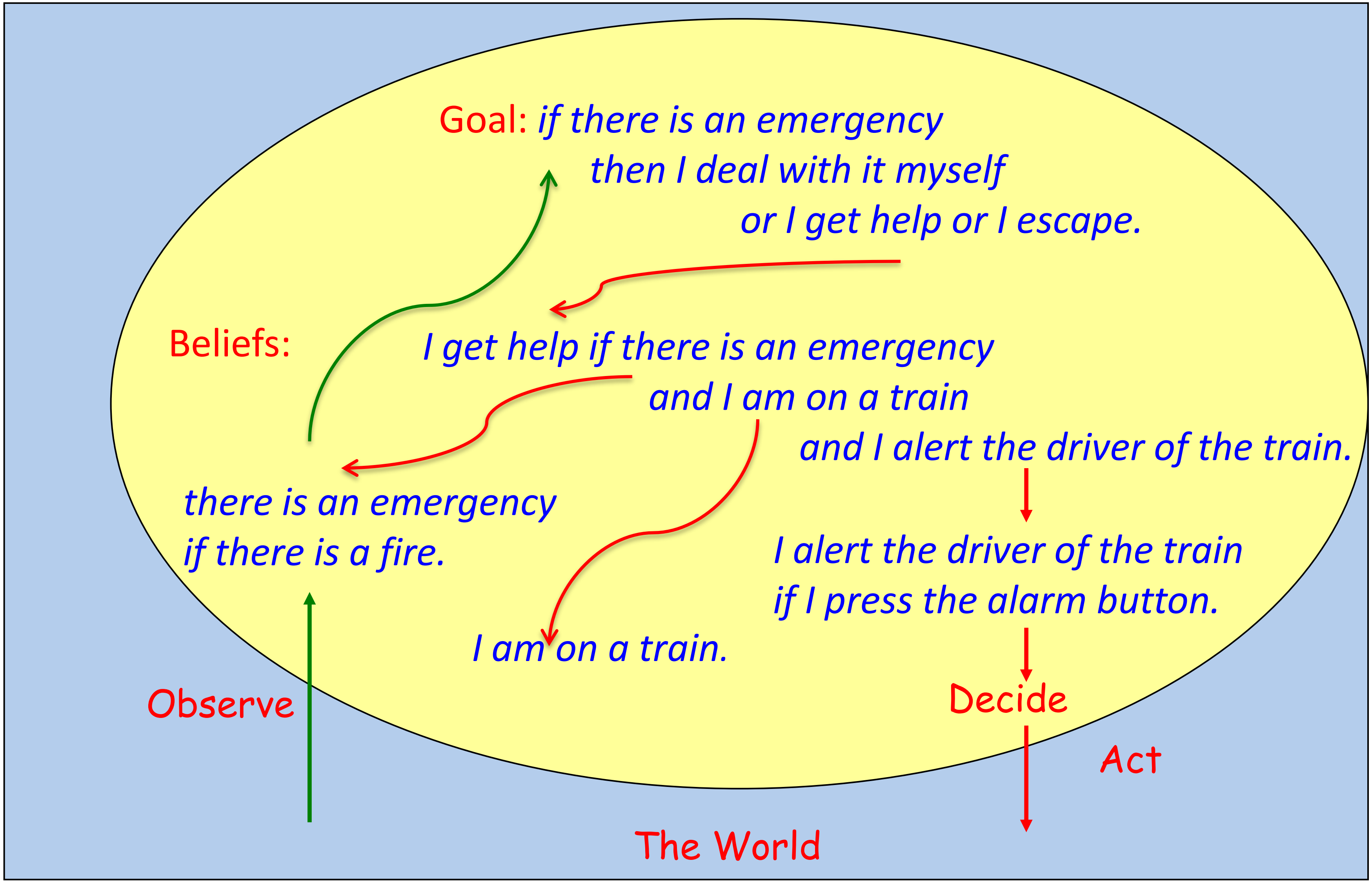
Artificial Intelligence and Human Thinking

- The Abductive Logic Programming (ALP) Agents
- ALP as the Language of Thought (LOT)
- ALP as a connectionist model of the mind

AI tools and techniques can be reconciled and combined



An Agent on the London Underground



Complex thinking and decision-making can be compiled into more efficient, lower-level maintenance goals, heuristics (or input-output associations)

For example:

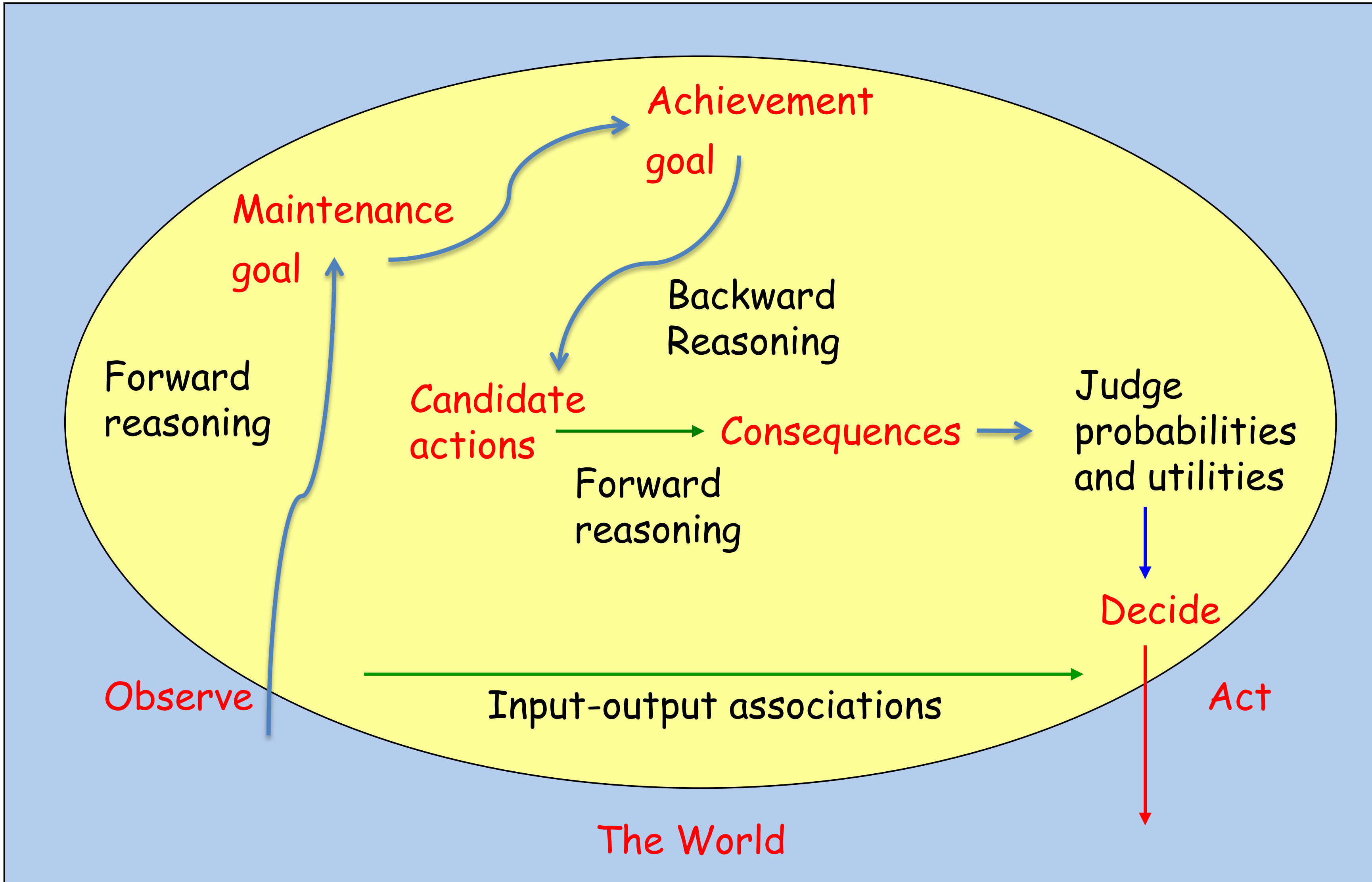
if there is a fire
and I am on a train
and I can not deal with the fire myself
then I press the alarm button.

Lower-level heuristics and higher-level thinking and deciding can be combined (as in dual process models of human thinking).

As Kahneman and Frederick (2002) put it:

> the intuitive, subconscious level “quickly proposes intuitive answers to judgement problems as they arise”, while the deliberative, conscious level “monitors the quality of these proposals, which it may endorse, correct, or override”.

Abductive Logic Programming (ALP) Agent



A Connectionist implementation of ALP

The syntax of beliefs in ALP agents

Beliefs have the form of logic programs:

conclusion
if $condition_1$ .... *and* $condition_n$

conclusion is an atomic formula
$condition_i$ are atomic formulas or negations of atomic formulas.

All variables are universally quantified.

The syntax of goals in ALP agents

Goals generalise FOL clauses:

If $condition_1$ *.... and* $condition_n$
then $conclusion_1$ *.... or* $conclusion_m$

$condition_i$ and $conclusion_i$ are atomic formulas.

All variables are universally quantified,
except variables in conclusions not in conditions,
which are existentially quantified.

ALP agents – minimal model semantics

Beliefs B describe the world as the agent sees it.
Goals G describe the world as the agent would like it to be.
Observations O describe the agent's perception of the world.

The agent's task is to generate a set Δ
of actions and assumptions about the world such that:

$G \cup O$ is true
in the minimal model of the world determined by
$B \cup \Delta$.

Truth is defined for arbitrary sentences of FOL.
Minimal models are defined by logic programs.

Goal *G*: *if there is an emergency*
then I deal with it myself
or I get help
or I escape

Observation *O*: *there is smoke*

Beliefs B: *there is smoke if there is a fire*
there is an emergency if there is a fire
I get help if I press the alarm button

$G \cup O$ is true in the minimal model of $B \cup \Delta$, where
Δ= {*there is a fire, I press the alarm button*}.

explains *there is smoke* — achieves *I get help*

Minimal model semantics

The agent's task is to generate Δ such that:

> $G \cup O$ is true
> in the minimal model of the world determined by
> $B \cup \Delta$.

The agent's world $B \cup \Delta$ is largely its own creation.

$B \cup \Delta$ needs to be consistent with
its observations O (i.e. O is true)
and make its goals G true.

The agent might be able to control its beliefs B,
goals G, actions and assumptions Δ,
but not its observations O.

Different Δ can solve the same task.

The challenge is to find the best Δ
within the computational resources available.

In classical decision theory, the value of an action
is measured by the expected utility of its consequences.

In philosophy of science, the value of an explanation
is measured by its probability and explanatory power.
(The more observations explained the better.)

ALP can associate probability with conditions

You will be rich	Consequence
if you buy a lottery ticket	Action
and your number is chosen.	State of the world = .0001

It will rain	Consequence
if you do a rain dance	Action
and the gods are pleased.	State of the world = ?

The ALP agent model can help people make better decisions.

In classical decision theory, all alternative actions and their consequences (outcomes) are given in advance.

In ALP agents, alternative actions and their consequences are generated by goals and beliefs.

The same criteria can be used both

to choose between alternatives and

to guide the search.

Artificial Intelligence and Human Thinking

- The Abductive Logic Programming (ALP) Agents
- ALP as the Language of Thought (LOT)
- ALP as a connectionist model of the mind

ALP as the Language of Thought (LOT)

In the philosophy of language, there are three schools of thought :

The LOT is a private, language-like representation, which is independent of public, natural language.

The LOT is a form of public, natural language. The natural languages that we speak influence the way we think.

Human thinking does not have a language-like structure at all.

In ALP agents, clausal logic serves as an agent's private LOT, independent of any public natural language.

ALP as the LOT

According to relevance theory [Sperber and Wilson, 1986],
people understand natural language by attempting
to extract the most information for the least processing cost.

It follows that:

If you want to identify the LOT,
then you should study communications
that are easy to understand.

If you want your communications to be easy to understand,
then you should express them in a form
that is close to the thoughts that you want to convey.

The Emergency Notice on the London underground

Press the alarm signal button to alert the driver.

The driver will stop
if any part of the train is in a station.

If not, the train will continue to the next station,
where help can more easily be given.

There is a 50 pound penalty for improper use.

The problems of understanding natural language communications

- Identify the intended meaning.

  he gave her the book.

- Represent the intended meaning in a canonical form.

  John gave Mary the book.
  John gave the book to Mary.
  Mary received the book from John.
  The book was given to Mary by John.

- Connect the canonical representation *e.g.*

  give(john, book, mary, e1000)

  to other goals and beliefs.

Clausal logic is a canonical form of FOL.

In clausal logic, sentences have a simplified form, e.g.:

has-feathers(X) ← bird(X).
bird(john).

In standard FOL, the same beliefs can be expressed in infinitely many, equivalent ways, including:

$$\neg(\exists X((\neg has\text{-}feathers\ (X) \wedge bird(X)) \vee \neg bird(john)))$$
$$\neg(\exists X((\neg has\text{-}feathers\ (X) \vee \neg bird(john)) \wedge\ (bird(X) \vee \neg bird(john))))$$

In clausal logic, reasoning can be reduced to forward or backward reasoning, which are special cases of the resolution rule.

ALP clausal logic, as a model of the LOT, can help people communicate

more clearly

by helping people to express communications
in a manner that is closer
to their logical form.

more simply

by helping people to express communications
in a form that is closer
to the canonical form of their meanings.

more coherently

by helping people to express communications
in a way that makes it easier
to link the meanings of new comunications to old information.

Artificial Intelligence and Human Thinking

- The Abductive Logic Programming (ALP) Agents
- ALP as the Language of Thought (LOT)
- ALP as a connectionist model of the mind

A Connectionist implementation of ALP

Goal: *if there is an emergency*
then I deal with it myself
or I get help or I escape.

Beliefs:

I get help if there is an emergency
and I am on a train
and I alert the driver of the train.

there is an emergency
if there is a fire.

I alert the driver of the train
if I press the alarm button.

I am on a train.

Connection graphs pre-compute links between the clauses, together with their unifying substitutions.

Links can be activated later
(by performing resolution)
when the need arises.

Any strategy can be used for activating links,
including forwards and backwards reasoning.

Links can be compiled into shortcuts,
which are like heuristic rules
and stimulus-response associations.

Connection graphs can combine logic, search, connectionism, learning and decision making

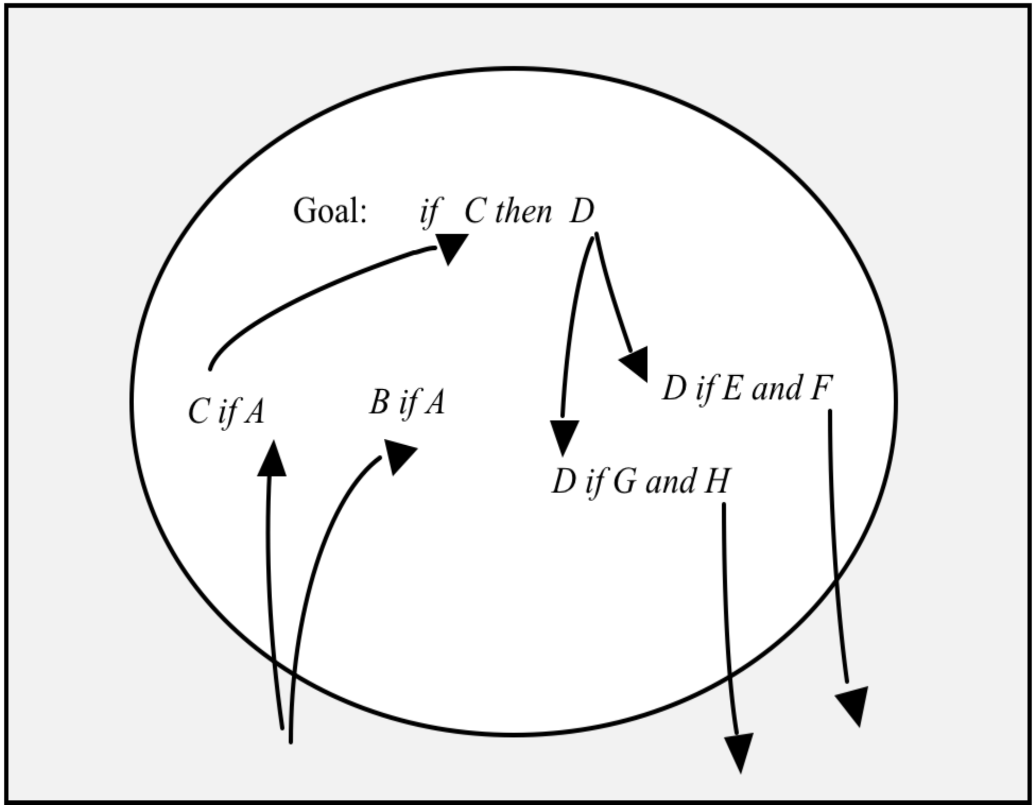
- Links can be weighted by statistics about how often they have contributed to successful outcomes in the past.

- Input observations and goals can be assigned different strengths (or utilities).

- The strength of observations and goals can be propagated in proportion to the weights on the links, as in the behavioural networks of Patie Maes.

- Activating links with the currently highest weighted strengths implements a form of best-first search for a solution with highest expected utility.

Internal clauses and links need not represent states of affairs in the real world



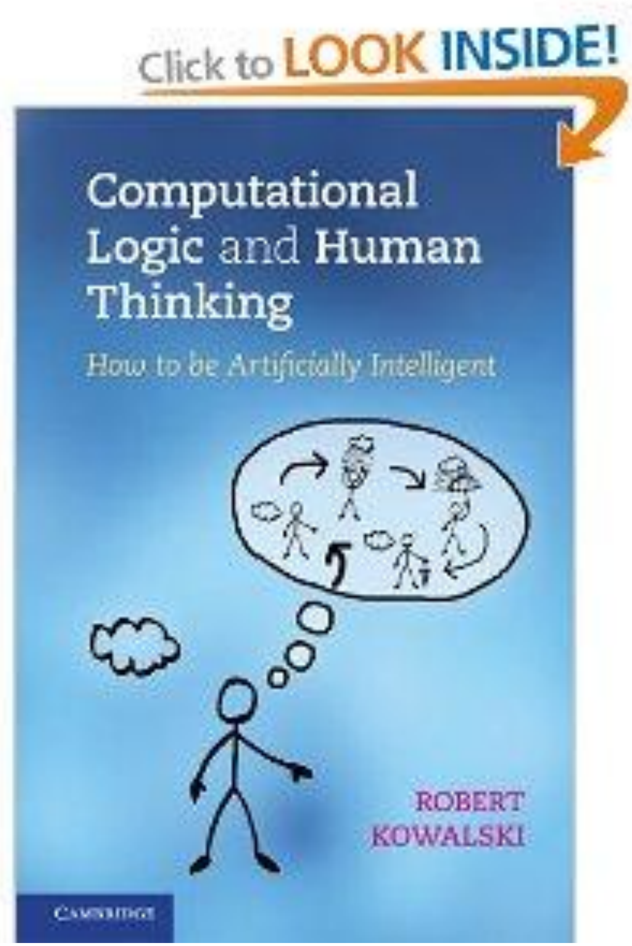
Only A, F and H correspond directly to reality.

C and D are mental constructs, which help to organise the agent's thoughts.

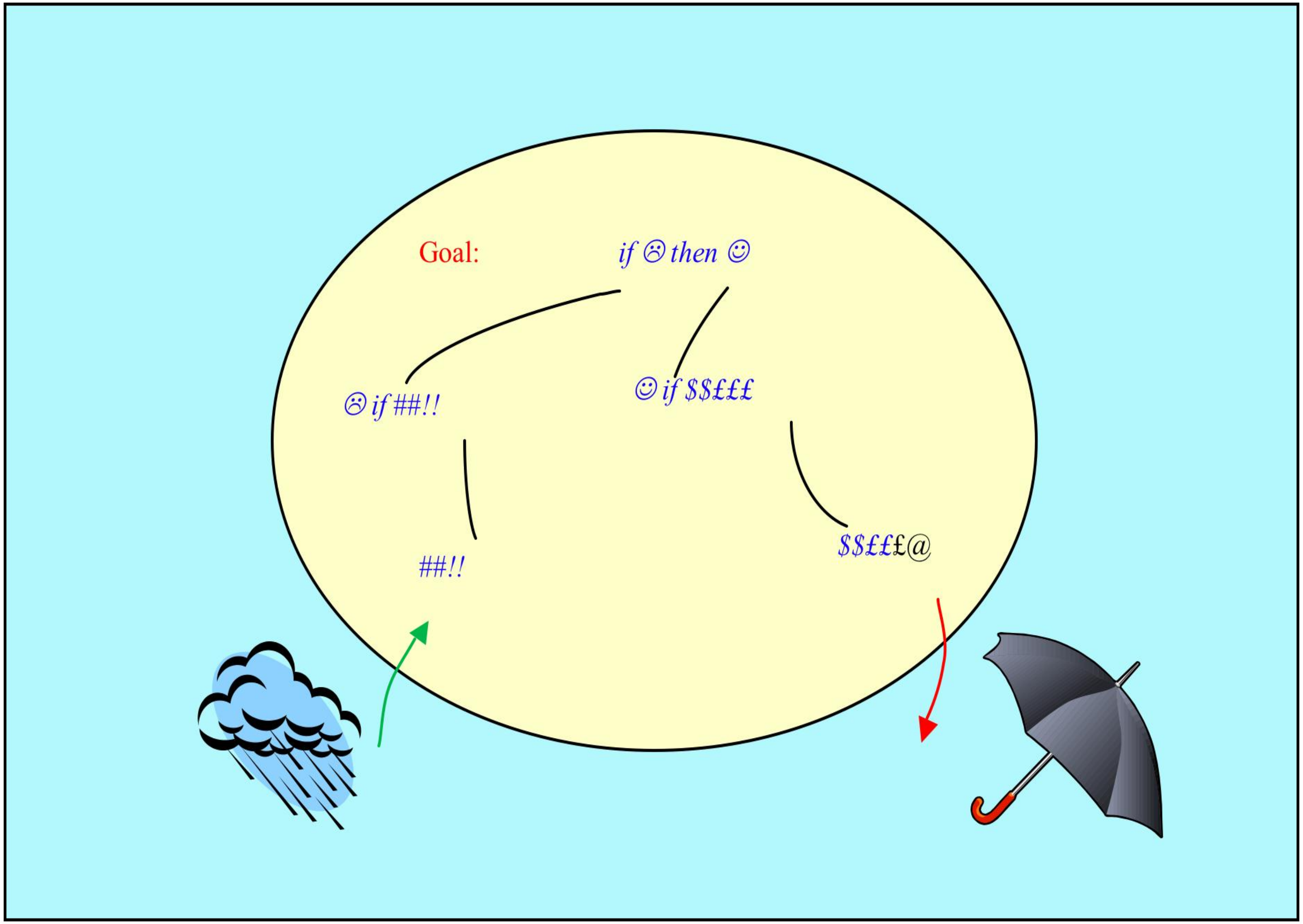
It can be difficult or impossible
to put our private thoughts into public words

Conclusions

- The ALP agent model can reconcile and combine the capabilities of
    - Logic
    - Procedural representations
    - Heuristics
    - Production Systems
    - Bayesian networks
    - Connectionism
- The ALP agent model, as a model of intelligent behaviour, can help people make better decisions.
- ALP clausal logic, as a model of the LOT, can help people communicate more effectively.


Click to LOOK INSIDE!
Computational Logic and Human Thinking
How to be Artificially Intelligent
ROBERT KOWALSKI
CAMBRIDGE

It can be difficult or impossible to put our private thoughts into public words



Standard FOL is to clausal form as natural language is to the LOT.

In both cases, inferences can be partitioned into two kinds:

The first kind converts sentences into canonical form.
The second kind reasons with that canonical form.

In FOL,
The first kind converts sentences into clausal form.
It includes Skolemization to eliminate existential quantifiers.

The second kind reasons with clausal form.
It includes deriving *P(t)* from $\forall X P(X)$ *),*
which is built into the resolution rule.

Natural Language and canonical form

The use of a canonical form makes it easier to use the information later.

In this case, the canonical form could be *give(john, mary, book)* or, say:

event(e1000)	*act(e1000, giving)*
agent(e1000, john)	*recipient(e1000, mary)*
object(e1000, book21)	*isa(book21, book)*

The Logic of the London Underground Notice

The first sentence

Press the alarm signal button to alert the driver.

is a procedural representation of a logic programming clause:

the driver is alerted
if you press the alarm signal button.

The Logic of the London Underground Notice

The second sentence

The driver will stop
if any part of the train is in a station.

is ambiguous, and one of its conditions has been omitted:

the driver will stop the train in a station
if the driver is alerted
and any part of the train is in the station.

The Logic of the London Underground Notice

The logic of the third sentence

If not, the train will continue to the next station,
where help can more easily be given.

is two sentences, say:

the driver will stop the train in the next station
if the driver is alerted
and not any part of the train is in a station.

help can more easily be given in an emergency
if the train is in a station.

The Logic of the London Underground Notice

The fourth sentence

There is a 50 pound penalty for improper use.

is also a conditional, but in disguise:

You may be liable to a £50 penalty
if you use the alarm signal button improperly.

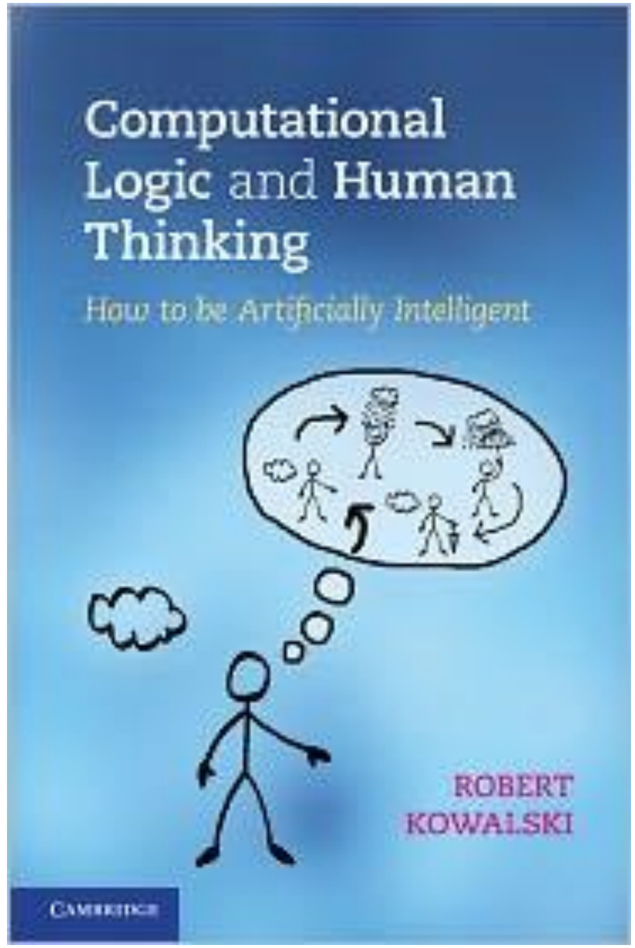
Different Δ can solve the same task.

The challenge is to find the best Δ
within the computational resources available.

In classical decision theory, the value of an action
is measured by the expected utility of its consequences.

In philosophy of science, the value of an explanation
is measured by its probability and explanatory power.
(The more observations explained the better.)

In ALP agents, the two measures of value can be combined.
Candidate actions and assumptions in Δ
can be evaluated by using forward reasoning
to generate their consequences.


Computational Logic and Human Thinking
How to be Artificially Intelligent
ROBERT KOWALSKI
CAMBRIDGE