AI: Present and Future

Chapter 27

Outline

- ♢ AI design problem
- ♢ Agent components
- ♢ Are we going in the right direction?
- ♢ What if AI does succeed?
- ♢ Conclusion

AI design problem

A unified view of AI as rational agent design

The design problem depends on

- the percepts and actions available to the agent
- the goals that the agent's behavior should satisfy
- the nature of the environment

AI design problem contd.

A variety of different agent designs are possible, ranging from
reflex agent _____ to _____ fully deliberative, knowledge-based agent

The components of these designs can have a number of different instantiations:

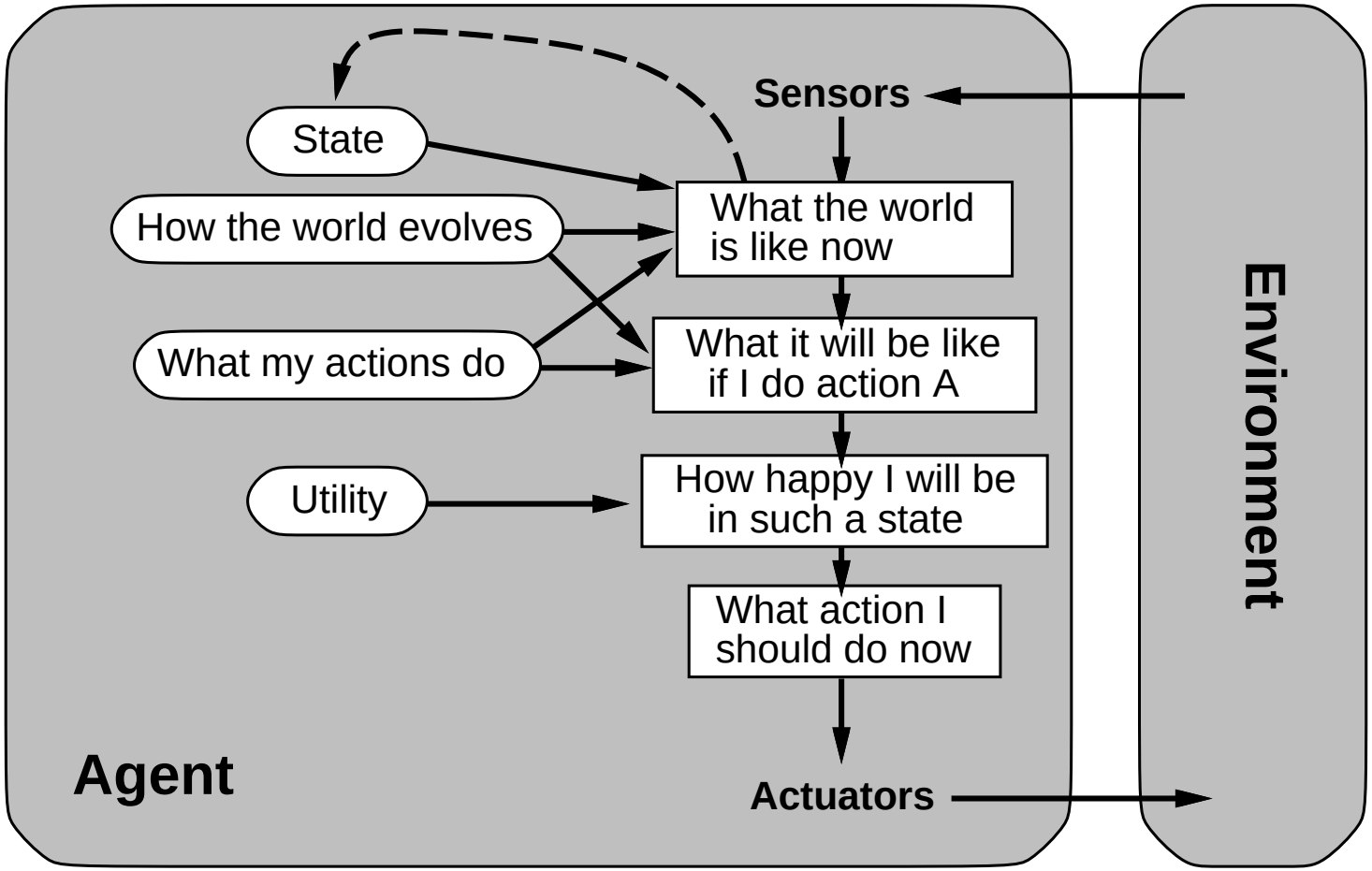
- ♢ logical
- ♢ probabilistic
- ♢ neural
- ♢ ...

For all the agent designs and components, there has been tremendous progress both in our scientific understanding and in our technological capabilities.

Will all this progress led to a general-purpose intelligent agent that can perform well in a wide variety of environments?

Agent Components

A model-based, utility-based agent



Agent Components contd.

- Interaction with the environment through sensors and actuators
- Keeping track of the state of the world
- Projecting, evaluating, and selecting future courses of action
- Utility as an expression of preferences
- Learning

Interaction with the environment

AI systems often were built in such a way that humans had to supply the inputs and interpret the outputs.

Robotic systems focused on low-level tasks in which high-level reasoning and planning were largely absent.

Keeping track of the state of the world

One of the core capabilities required for an intelligent agent.

Requires both:

- ♢ Perception
- ♢ Updating of internal representations

Using:

- ♢ propositional logic techniques
- ♢ first-order logic techniques
- ♢ filtering algorithms for tracking uncertain environments.

Objects in an uncertain environment: problem of identity uncertainty:
We don't know which object is which.

this problem has been largely ignored in logic-based AI,
where it has generally been assumed that percepts incorporate symbols
that identify the objects.

Selecting future courses of action

Imposing hierarchical structure on behavior.

Utility as an expression of preferences

Realistic utility functions

Learning

Inductive learning (supervised, unsupervised, reinforcement-based)

Machine learning has made very little progress on the important problem of constructing new representations at levels of abstraction higher that the input vocabulary

e.g. how can an agent generate useful predicates such as $Office$ or $Cafe$

if they are not supplied to it by humans?

♢ success and understanding decreases with representation complexity

♢ use of background knowledge is poor

Agent architectures

Which of the agent architectures should an agent use???

Agent architectures

Which of the agent architectures should an agent use??

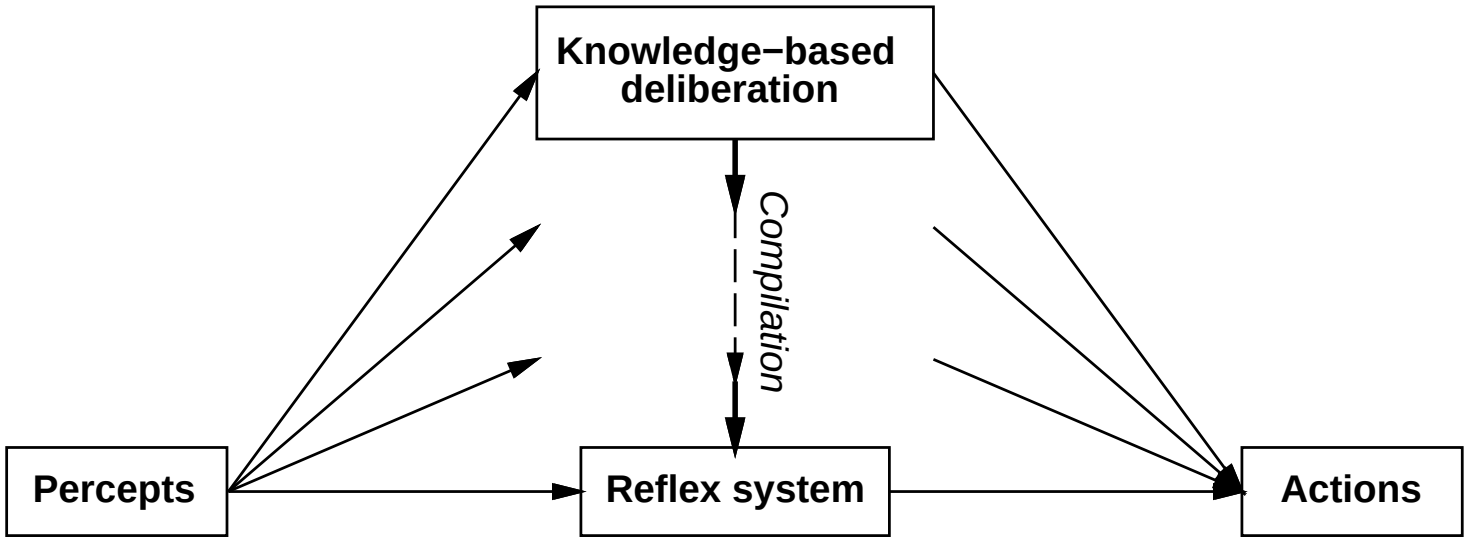
Answer: All of them! (reflex, model-based, goal-based, utility-based)

Reflex responses are needed for situations in which time is of the essence.

Knowledge-based deliberation allows the agent to plan ahead

A complete agent must be able to do both, using a hybrid architecture

Hybrid architecture



the boundaries between different decision components are not fixed.

compilation serves to convert deliberative decision making into more efficient, reflexive mechanisms.

compilation generalizes deliberation into reflex.

– Agent architectures such as SOAR and THEO

Real-time AI

Agents also need ways to control their own deliberation.

As AI systems move into more complex domains, all problems will become real-time, because the agent will never have long enough to solve the decision problem exactly.

Methods that work in more general decision-making situations:

♢ Anytime algorithms
- solution improves gradually over time
- but always available
- e.g., iterative deepening

♢ Decision-theoretic meta-reasoning
- decide what computation to perform
- based on cost versus benefit tradeoff

Real-time AI contd.

As AI systems move into more complex domains, all problems will become real-time, because the agent will never have long enough to solve the decision problem exactly.

Methods that work in more general decision-making situations:

- Anytime algorithms
  an algorithm whose output quality improves gradually over time, so that it has a reasonable decision ready whenever it is interrupted.
  – a simple example: IDS in game playing
- Decision-theoretic metareasoning
  this method applies the theory of information value to the selection of computations.

  Metareasoning is one aspect of a general reflective architecture:
  an architecture that enables deliberation about the computational entities and actions occurring within the architecture itself.

Are We Going in the Right Direction?

AI's current path is more like a tree climb or a rocket trip?

Our goal was to build agents that act rationally. but . . .

Achieving perfect rationality – always doing the right thing – is not feasible in complicated environments. The computational demands are just too high. However, hypothesis of perfect rationality is a good starting point for analysis.

What exactly the goal of AI is??

- ♢ Perfect rationality
- ♢ Calculative rationality
- ♢ Bounded rationality
- ♢ Bounded optimality

Perfect rationality

A perfectly rational agent acts at every instant in such a way as to maximize its expected utility, given the information it has acquired from the environment.

♢ classical decision theory
♢ maximize expected utility

– computationally infeasible for complex environments

Calculative rationality

Calculative rationality, is the notion of rationality that we have used implicitly in designing logical and decision-theoretic agents.

A calculatively rational agent eventually returns what would have been the rational choice at the beginning of its deliberation.

But, in most environments, the right answer at the wrong time is of no value.

♢ will eventually produce a rational choice
  – as defined by some calculation
  – may take a long time

Bounded rationality

Herbert Simon (1957) rejected the notion of perfect (or even approximately perfect) rationality and replaced it with bounded rationality.

Bounded rationality work primarily by satisficing: deliberating only long enough to come up with that answer that is good enough.

Simon won the Nobel prize in economics for this work.

It appears to be a useful model of human behaviors in many cases.

Bounded optimality (BO)

A bounded optimal agent behaves as well as possible given its computational resources

- ♢ expected utility no worse than any other agent under same constraints
- ♢ most feasible choice for intelligent agent rationality

Of these four possibilities, bounded optimality seems to offer the best hope for a strong theoretical foundation for AI.

It has the advantage of being possible to achieve: there is always at least on best program.

What if AI does succeed?

We can expect that medium-level successes in AI would affect all kind of people in their daily lives.

A technological capability at this level might also be applied to the development of autonomous weapons!

It seems likely that a large-scale success in AI would change the lives of a majority of humankind. At this level, AI systems could pose a more direct threat to human autonomy, freedom, and even survival!

$\Rightarrow$

We cannot divorce AI research from its ethical consequences.

Conclusion

We see that AI has made great progress in its short history, but the final sentence of Alan Turing's essay on *Computing Machinery and Intelligence* is still valid today:

We can see only a short distance ahead, but we can see that much remains to be done.