Chapter 1: Intelligent Agents

What is Artificial Intelligence?

- Artificial Intelligence is the synthesis and analysis of computational agents that act intelligently.
- An agent is something that acts in an environment.
- An agent acts intelligently if:
  - its actions are appropriate for its goals and circumstances
  - it is flexible to changing environments and goals
  - it learns from experience
  - it makes appropriate choices given perceptual and computational limitations

Observing and investigating intelligence

- reflexes
- insticts
- problem solving
- learning
- tool use
- intellectual tasks
- language communication
- creativity
- self-recognition

Reflexes

involuntary and nearly instantaneous movement in response to a stimulus

can be innate or aquired (conditioning)

Reflexes

- Can intelligence emerge from the interplay of reflexes?
- Is building structure always a sign of intelligent behaviour?

Instincs

Migrating animals: birds, butterflies, mammals, ...

Instincs

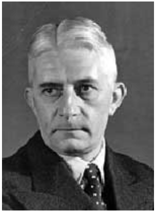
the inherent inclination of a living organism toward a particular complex behavior.

- innate behavior: absence of learning
- eating, nest building, mating, feeding, migration, ...
- no intelligent behaviour, can be counterproductive
- can be overridden by competing instincts or reasoning

Problem solving

Wolfgang Köhler (1887 – 1967)

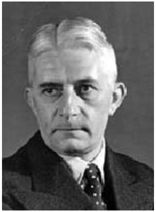
- Intelligenzprüfungen an Anthropoiden (1917)
- The Mentality of Apes (1925)



Problem solving

Wolfgang Köhler (1887 – 1967)

Intelligenzprüfungen an Anthropoiden (1917)
The Mentality of Apes (1925)



Problem solving

- behaviour is driven by insight, if a goal cannot be reached on a direct and simple path, but a way round is available and chosen.
- distinction between true intelligent results and coincidental success.
  - chance: several independent movements
  - true intelligence: spatially and temporally coherent course without hesitation

Problem solving

Robert Epstein (1953 –)

are pigeons as intelligent as chimpanzees?

Problem solving

Bernd Heinrich (1953 –)

problem solving in ravens

Ethology **111,** 962—976 (2005)
© 2005 Blackwell Verlag, Berlin

Testing Problem Solving in Ravens: String-Pulling to Reach Food

Bernd Heinrich & Thomas Bugnyar

Department of Biology, University of Vermont, Burlington, VT, USA

Abstract

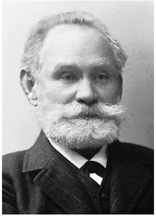
The aim of our study was to re-examine the acquisition of problem-solving behaviour in ravens: accessing meat suspended from a perch by a string. In contrast to a previous study, here we: (i) controlled for possible effects of fear of the string, competition by dominants, and social learning and (ii) devised a mechanically equivalent but non-intuitive task to test for the possibility of means–end understanding. One-year-old ravens confronted with meat on a string for the first time tried several ways to reach the food. However, five of six birds suddenly performed a coherent sequence of pulling up and stepping on loops of string, essential for solving the problem. Those five birds were also successful in the non-

Problem solving

Making the task more challenging ...

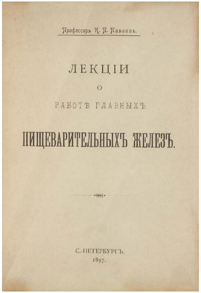
Animals mastered the second task only after they have solved the first one!

Conditioning



Iwan Petrowitsch Pawlow (1849 – 1936)

- acquired/learned reflexes
- association learning: direct linking of sensual stimuli
- assumed to be the only type of learning in animals



Профессоръ И. П. Павловъ.

ЛЕКЦІИ

О

РАБОТѢ ГЛАВНЫХЪ

ПИЩЕВАРИТЕЛЬНЫХЪ ЖЕЛЕЗЪ.

С.-ПЕТЕРБУРГЪ.
1897.

Imprinting

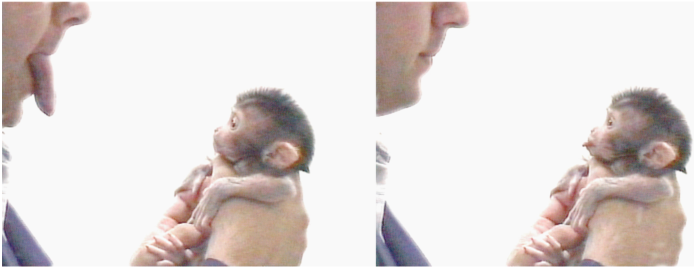
phase-sensitive learning

- rapid
- apparently independent of the consequences of the behavior

Imitation

observing and replicating another individuums behavior

- form of social learning
- leads to the development of traditions and culture
- transfer of information (behaviours, customs, etc.) across generations without the need for genetic inheritance



Imitation

blue tits learned to open sealed milk bottles

1900 – 1945	open milk bottles were dropped at the doorstep
around 1945	sealed milk bottles have been introduced
since 1960	doorstep delivery has been suspended

around 1945	first indiviuals learned to open the seal
around 1950	the whole population of British blue tits had aquired the skill

Imitation

Are blue tits smarter than european robins?

- Only few individuals acquired the skill but not the whole population. Why?

Imitation

Are blue tits smarter than ravens?

Imitation

Are blue tits smarter than ravens?

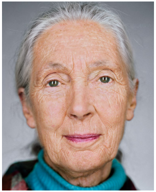
The string pulling experiment again:

- Every individual developed its own solution strategy
- No transfer of skills between individuals
- One raven didn't master the task even though it was able to observe the other ones pulling the meat

Tool use

Jane Goodall (1934 –)

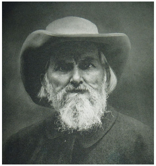
Tool-Using and Aimed Throwing in a Community of Free-Living Chimpanzees (1964)

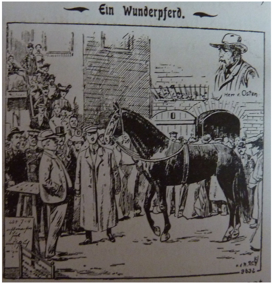


Intellectual tasks

Der kluge Hans (1895 – 1916 ?)

- Orlow-Traber
- owner: Wilhelm von Osten (1838 – 1909)
- ability to count and calculate
  and to distinguish simple concepts





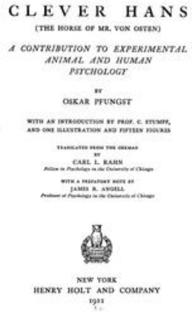
Intellectual tasks

Has Hans been intelligent?

- Hans commission (1904): no manipulation found

Intellectual tasks

Oskar Pfungst (1874 – 1933)


CLEVER HANS
(THE HORSE OF MR. VON OSTEN)

A CONTRIBUTION TO EXPERIMENTAL ANIMAL AND HUMAN PSYCHOLOGY

BY
OSKAR PFUNGST

WITH AN INTRODUCTION BY PROF. C. STUMPF, AND ONE ILLUSTRATION AND FIFTEEN FIGURES

TRANSLATED FROM THE GERMAN
BY
CARL L. RAHN

WITH A PREFATORY NOTE BY
JAMES R. ANGELL

NEW YORK
HENRY HOLT AND COMPANY
1911

Rigorous testing:

- Isolating horse and questioner from spectators
- Using questioners other than the horse's master
- Blindfolding the horse
- Using questions, where the questioner didn't know the answer

Intellectual tasks

- Hans could get the correct answer even if not von Osten was asking

Intellectual tasks

- Hans could get the correct answer even if not von Osten was asking
- horse must have seen the questioner

Intellectual tasks

- Hans could get the correct answer even if not von Osten was asking
- horse must have seen the questioner
- questioner had to know what the answer was

Intellectual tasks

- Hans could get the correct answer even if not von Osten was asking
- horse must have seen the questioner
- questioner had to know what the answer was
- von Osten knew the answer $\rightarrow$ 89% correct
- von Osten did not know the answer $\rightarrow$ only 6% correct

Intellectual tasks

- Hans could get the correct answer even if not von Osten was asking
- horse must have seen the questioner
- questioner had to know what the answer was
- von Osten knew the answer $\rightarrow$ 89% correct
- von Osten did not know the answer $\rightarrow$ only 6% correct
- the questioner's posture and facial expression changed as the horse's taps approached the right answer
- the changes were consistent with an increase in tension
- the tension was released when the horse made the final, correct tap

Intellectual tasks

Confirming the findings by inverse experiments:

- cues from the audience are sufficiently reliable

$\Rightarrow$ the Clever Hans effect

Intellectual tasks

Irene Pepperberg (1949 –)

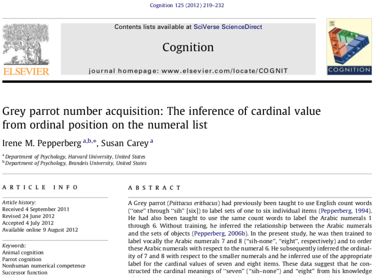
Teaching numerical concepts to parrots

Intellectual tasks

Irene Pepperberg (1949 –)

Teaching numerical concepts to parrots

- counting (up to seven)
- simple inferences on numbers
- cardinal vs. ordinal numbers
- the notion of zero



Cognition 125 (2012) 219–232

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Grey parrot number acquisition: The inference of cardinal value from ordinal position on the numeral list

Irene M. Pepperberg [a,b,*], Susan Carey [a]

[a] Department of Psychology, Harvard University, United States
[b] Department of Psychology, Brandeis University, United States

ARTICLE INFO

Article history:
Received 4 September 2011
Revised 24 June 2012
Accepted 4 July 2012
Available online 9 August 2012

Keywords:
Animal cognition
Parrot cognition
Nonhuman numerical competence
Successor function

ABSTRACT

A Grey parrot (*Psittacus erithacus*) had previously been taught to use English count words (''one'' through ''sih'' [six]) to label sets of one to six individual items (Pepperberg, 1994). He had also been taught to use the same count words to label the Arabic numerals 1 through 6. Without training, he inferred the relationship between the Arabic numerals and the sets of objects (Pepperberg, 2006b). In the present study, he was then trained to label vocally the Arabic numerals 7 and 8 (“sih-none”, “eight”, respectively) and to order these Arabic numerals with respect to the numeral 6. He subsequently inferred the ordinality of 7 and 8 with respect to the smaller numerals and he inferred use of the appropriate label for the cardinal values of seven and eight items. These data suggest that he constructed the cardinal meanings of ''seven'' (“sih–none”) and “eight” from his knowledge

Natural language communication

- Alex the grey parrot
- Washoe the chimp
- Koko the gorilla
- Nim Chimpsky

Alex the grey parrot

Irene Pepperberg (1949 –)
Alex (1976 – 2007)

Alex the grey parrot

MAN	Come on, what is it?
ALEX	Keychain.
I.P.	Good birdie. Good parrot. What is it?
ALEX	Rock.
MAN	Good boy.
I.P.	Yeah, good birdie. Alex, what toy?
ALEX	Nail.
I.P.	Nail, that's right. You're a good birdie. You're a very good boy.
MAN	What toy?
ALEX	Truck.
I.P.	That's right.
MAN	You're a very good birdie.
I.P.	Tell me what color. What color?
ALEX	Yellow.
I.P.	Yellow, that's right.
MAN	What matter?
ALEX	Wood.
MAN	Good. That's right. Very good.
I.P.	How many? Good boy. How many?
ALEX	Two.
I.P.	Good parrot. Good boy. One. Two.

Alex the grey parrot

I.P.	Can you tell me what's different? What's different?
ALEX	Color.
I.P.	Good boy. All right. What same? What same?
ALEX	Shape.
I.P.	Good boy, good birdie. What color bigger? You know. What color bigger?
ALEX	Yellow.
I.P.	Good boy. Good birdie.
I.P.	Look. What matter four-corner blue?
DENISE	What matter four-corner blue?
ALEX	Wood.
DENISE	That's a good boy. You're right.

Alex the grey parrot

After 19 years of training:

- 200/500 lexical items (active/passive)
- basic language understanding and language production capabilities
- different dimensions of object descriptions (color, shape, material, ...)
- complex object descriptions: *four-corner wood*
- refusal to cooperate; rejection of food or toys

Washoe the chimp

Allen Gardner (1930 –)
Beatrice Gardner (1934 – 1995)
Washoe (1965 – 2007)
Teaching American Sign Language

- raised like a child
- private trailer with living and cooking areas
- learned 350 words of ASL
- taught her adopted son Loulis some ASL
- invention of new sign combinations:
  swan = *water* + *bird*
- simple verb-noun combinations
- no conditioning, no rewards

Washoe the chimp

Allen Gardner (1930 –)
Beatrice Gardner (1934 – 1995)
Washoe (1965 – 2007)
Teaching American Sign Language

- raised like a child
- private trailer with living and cooking areas
- learned 350 words of ASL
- taught her adopted son Loulis some ASL
- invention of new sign combinations:
  $swan = water + bird$
- simple verb-noun combinations
- no conditioning, no rewards

Koko the gorilla

Francine (Penny) Patterson (1947 –)
Ronald Cohn (?)
Koko (1971 –)

- production: 1000 signs, comprehension: 2000 signs
- learning or conditioning?
- invention of new sign combinations: *ring* = *finger* + *bracelet*
- no sentences, but adjectives, nouns and noun phrases
- rewards if answer is somehow ok: *apple* is *apple* or *red*

Nim Chimpsky

Herbert S. Terrace

Nim Chimpsky (1973 – 2000)

- not "raised like a child"
- learned 150 signs
- but Terrace concluded that this was not natural language
    - mean length of 20000 recorded responses only 1.2
    - no correlation between lexical growth and structural complexity

Nim Chimpsky

strong criticism of the experimental methodology

- apes remained passive
- interpretation of responses is up to the experimenter
- human language capabilities are projected onto the ape
- often the response of the ape was preceded by an (unvoluntary) similar movement of the experimenter (250 ms earlier)

Creativity

Music playing robots

Creativity

Painting elephants

Creativity

Desmond Morris (1928 –)
Painting chimpanzees

Self-recognition

Does the animal recognize itself as an acting agent?
$\Rightarrow$ mirror test

Passing the test:

Humans (at age 18 months or older), Bonobos, Chimpanzees, Orangutans, Gorillas (Koko!), Bottlenose dolphins, Orcas, Elephants, European Magpies (the only non-mammal)

Conclusions

- intelligence is a composite property
  - different kinds of intelligence

Conclusions

- intelligence is a composite property
  - different kinds of intelligence
- intelligence is a graded notion
  - cognitive tasks with different degrees of difficulty

Conclusions

- intelligence is a composite property
  - different kinds of intelligence
- intelligence is a graded notion
  - cognitive tasks with different degrees of difficulty
- intelligence is a highly subjective notion
  - contradictory interpretations of observations are quite common

Conclusions

- intelligence is a composite property
  - different kinds of intelligence
- intelligence is a graded notion
  - cognitive tasks with different degrees of difficulty
- intelligence is a highly subjective notion
  - contradictory interpretations of observations are quite common
- danger of "anthropo" morphization
  - objective investigation methods required

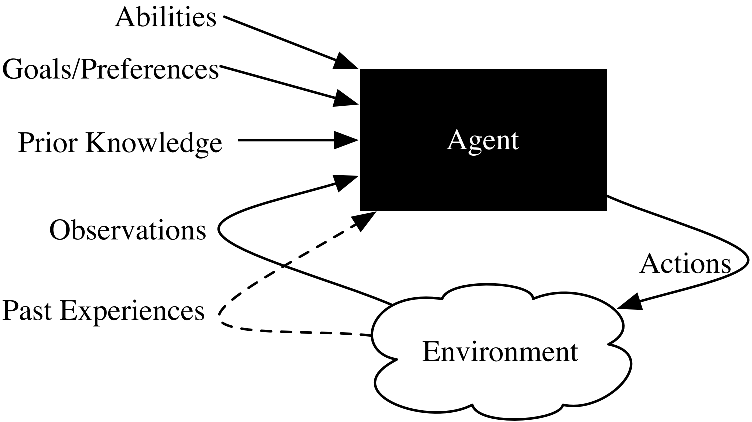
Conclusions

- intelligence is a composite property
  - different kinds of intelligence
- intelligence is a graded notion
  - cognitive tasks with different degrees of difficulty
- intelligence is a highly subjective notion
  - contradictory interpretations of observations are quite common
- danger of "anthropo" morphization
  - objective investigation methods required
- studying intelligence requires systematic analysis
  - no conclusions can be drawn from isolated observations

Goals of Artificial Intelligence

- Scientific goal: to understand the principles that make intelligent behavior possible in natural or artificial systems.
  - analyze natural and artificial agents
  - formulate and test hypotheses about what it takes to construct intelligent agents
  - design, build, and experiment with computational systems that perform tasks that require intelligence
- Engineering goal: design useful, intelligent artifacts.
- Analogy between studying flying machines and thinking machines.

Agents acting in an environment



Examples of Agents

- Organisations Microsoft, Al Qaeda, Government of Canada, UBC, CS Dept,...
- People teachers, physicians, stock traders, engineers, researchers, travel agents, farmers, waiters...
- Computers/devices thermostats, user interfaces, airplane controllers, network controllers, games, advising systems, tutoring systems, diagnostic assistants, robots, Google car, Mars rover...
- Animals dogs, mice, birds, insects, worms, bacteria...

Inputs to an agent

- Abilities — the set of things it can do
- Goals/Preferences — what it wants, its desires, its values,...
- Prior Knowledge — what it comes into being knowing, what it doesn't get from experience,...
- History of observations (percepts, stimuli) of the environment
  - (current) observations — what it observes now
  - past experiences — what it has observed in the past

Example agent: robot

- abilities: movement, grippers, speech, facial expressions,...
- goals: deliver food, rescue people, score goals, explore,...
- prior knowledge: what is an important feature, which categories of objects can be distinguished, what a sensor can tell us,...
- observations: vision, sonar, sound, speech recognition, gesture recognition,...
- past experiences: effect of steering, slipperiness, how people move,...

Example agent: teacher

- abilities: present new concept, drill, give test, explain concept,. . .
- goals: particular knowledge, skills, inquisitiveness, social skills,. . .
- prior knowledge: subject material, teaching strategies,. . .
- observations: test results, facial expressions, errors, focus,. . .
- past experiences: prior test results, effects of teaching strategies, . . .

Dimensions of Complexity

- Research proceeds by making simplifying assumptions, and gradually reducing them.
- Each simplifying assumption gives a dimension of complexity
  - multiple values in a dimension: from simple to complex
  - simplifying assumptions can be relaxed in various combinations

Dimensions of Complexity

- Flat or modular or hierarchical
- Explicit states or features or individuals and relations
- Static or finite stage or indefinite stage or infinite stage
- Fully observable or partially observable
- Deterministic or stochastic dynamics
- Goals or complex preferences
- Single-agent or multiple agents
- Knowledge is given or knowledge is learned from experience
- Perfect rationality or bounded rationality

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to see the Mona Lisa in Paris.

Modularity

- Model at one level of abstraction: flat
- Model with interacting modules that can be understood separately: modular
- Model with modules that are (recursively) decomposed into modules: hierarchical
- Example: Planning a trip from here to see the Mona Lisa in Paris.
- Flat representations are adequate for simple systems.
- Complex biological systems, computer systems, organizations are all hierarchical
- A flat description is either continuous or discrete. Hierarchical reasoning is often a hybrid of continuous and discrete.

Succinctness and Expressiveness

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.

A agent can reason in terms of:

- Explicit states — a state is one way the world could be

Succinctness and Expressiveness

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.

A agent can reason in terms of:

- Explicit states — a state is one way the world could be
- Features or propositions.
  - States can be described using features.
  - 30 binary features can represent $2^{30} = 1,073,741,824$ states.

Succinctness and Expressiveness

Much of modern AI is about finding compact representations and exploiting the compactness for computational gains.
A agent can reason in terms of:

- Explicit states — a state is one way the world could be
- Features or propositions.
  - States can be described using features.
  - 30 binary features can represent $2^{30} = 1,073,741,824$ states.
- Individuals and relations
  - There is a feature for each relationship on each tuple of individuals.
  - Often an agent can reason without knowing the individuals or when there are infinitely many individuals.

Planning horizon

...how far the agent looks into the future when deciding what to do.

- Static: world does not change
- Finite stage: agent reasons about a fixed finite number of time steps
- Indefinite stage: agent reasons about a finite, but not predetermined, number of time steps
- Infinite stage: the agent plans for going on forever (process oriented)

Uncertainty

There are two dimensions for uncertainty

- uncertain dynamics
- uncertain perception (sensor information and its interpretation)

Uncertainty

There are two dimensions for uncertainty

- uncertain dynamics
- uncertain perception (sensor information and its interpretation)

In each dimension an agent can have

- No uncertainty: the agent knows which world is true
- Disjunctive uncertainty: there is a set of worlds that are possible
- Probabilistic uncertainty: a probability distribution over the possible worlds.

Why Probability?

- Agents need to act even if they are uncertain.
- Predictions are needed to decide what to do:
  - definitive predictions: you will be run over tomorrow
  - disjunctions: be careful or you will be run over
  - point probabilities: probability you will be run over tomorrow is 0.002 if you are careful and 0.05 if you are not careful
  - probability ranges: you will be run over with probability in range [0.001,0.34]
- Acting is gambling: agents who don't use probabilities will lose to those who do.
- Probabilities can be learned from data and prior knowledge.

Uncertain dynamics

If an agent knew the initial state and its action, could it predict the resulting state?
The dynamics can be:

- Deterministic : the resulting state is determined from the action and the state
- Stochastic : there is uncertainty about the resulting state.

Perceptual Uncertainty

Whether an agent can determine the state from its observations:

- Fully-observable: the agent can observe the state of the world.
- Partially-observable: there can be a number states that are possible given the agent's observations.

Goals or complex preferences

- achievement goal is a goal to achieve. This can be a complex logical formula.
- complex preferences may involve tradeoffs between various desiderata, perhaps at different times.
  - ordinal only the order matters
  - cardinal absolute values also matter
- Examples: coffee delivery robot, medical doctor

Single agent or multiple agents

- **Single agent** reasoning is where an agent assumes that any other agents are part of the environment.
- **Multiple agent** reasoning is when an agent reasons strategically about the reasoning of other agents.

Agents can have their own goals: cooperative, competitive, or goals can be independent of each other

Learning from experience

Whether the model is fully specified a priori:

- Knowledge is given.
- Knowledge is learned from data or past experience.

Perfect rationality or bounded rationality

- Perfect rationality: the agent can determine the best course of action, without taking into account its limited computational resources.
- Bounded rationality: the agent must make good decisions based on its perceptual, computational and memory limitations.

Dimensions of Complexity

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single-agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

State-space Search

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Classical Planning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Decision Networks

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Markov Decision Processes (MDPs)

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Decision-theoretic Planning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Reinforcement Learning

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Classical Game Theory

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

Humans

- flat or modular or hierarchical
- explicit states or features or individuals and relations
- static or finite stage or indefinite stage or infinite stage
- fully observable or partially observable
- deterministic or stochastic dynamics
- goals or complex preferences
- single agent or multiple agents
- knowledge is given or knowledge is learned
- perfect rationality or bounded rationality

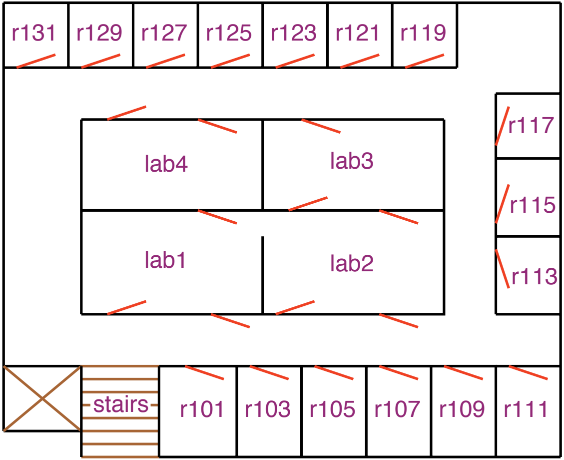
The Dimensions Interact in Complex Ways

- Partial observability makes multi-agent and indefinite horizon reasoning more complex
- Modularity interacts with uncertainty and succinctness: some levels may be fully observable, some may be partially observable
- Three values of dimensions promise to make reasoning simpler for the agent:
  - Hierarchical reasoning
  - Individuals and relations
  - Bounded rationality

Four Example Application Domains

- Autonomous delivery robot roams around an office environment and delivers coffee, parcels,...
- Diagnostic assistant helps a human troubleshoot problems and suggests repairs or treatments. E.g., electrical problems, medical diagnosis.
- Intelligent tutoring system teaches students in some subject area.
- Trading agent buys goods and services on your behalf.

Domain for Delivery Robot



Autonomous Delivery Robot

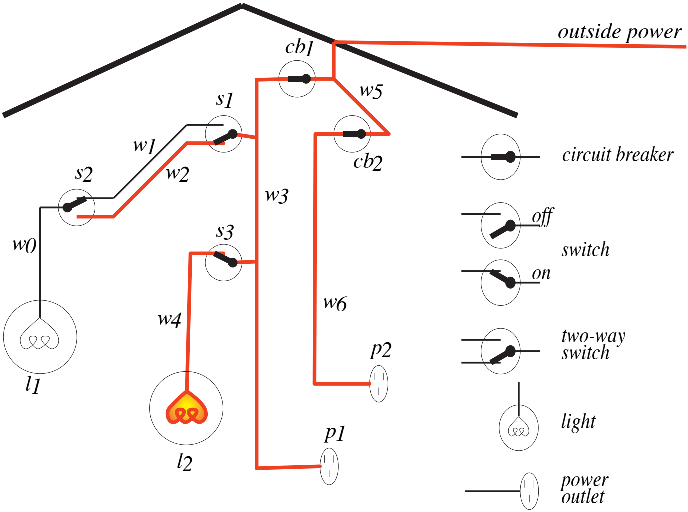
Example inputs:

- Abilities: movement, speech, pickup and place objects.
- Prior knowledge: its capabilities, objects it may encounter, maps.
- Past experience: which actions are useful and when, what objects are there, how its actions affect its position.
- Goals: what it needs to deliver and when, tradeoffs between acting quickly and acting safely.
- Observations: about its environment from cameras, sonar, sound, laser range finders, or keyboards.

What does the Delivery Robot need to do?

- Determine where Craig's office is. Where coffee is. . .
- Find a path between locations.
- Plan how to carry out multiple tasks.
- Make default assumptions about where Craig is.
- Make tradeoffs under uncertainty: should it go near the stairs?
- Learn from experience.
- Sense the world, avoid obstacles, pickup and put down coffee.

Domain for Diagnostic Assistant



Diagnostic Assistant

Example inputs:

- Abilities: recommends fixes, ask questions.
- Prior knowledge: how switches and lights work, how malfunctions manifest themselves, what information tests provide, the side effects of repairs.
- Past experience: the effects of repairs or treatments, the prevalence of faults or diseases.
- Goals: fixing the device and tradeoffs between fixing or replacing different components.
- Observations: symptoms of a device or patient.

Subtasks for the diagnostic assistant

- Derive the effects of faults and interventions.
- Search through the space of possible fault complexes.
- Explain its reasoning to the human who is using it.
- Derive possible causes for symptoms; rule out other causes.
- Plan courses of tests and treatments to address the problems.
- Reason about the uncertainties/ambiguities given symptoms.
- Trade off alternate courses of action.
- Learn what symptoms are associated with faults, the effects of treatments, and the accuracy of tests.

Trading Agent

Trading agent interacts with an information environment to purchase goods and services.

- It acquires a users needs, desires and preferences.
- It finds what is available.
- It purchases goods and services that fit together to fulfill your preferences.
- It is difficult because users preferences and what is available can change dynamically, and some items may be useless without other items.

Trading Agent Inputs

- **Abilities:** acquire information, make recommendations, purchase items.
- **Prior knowledge:** the ontology of what things are available, where to purchase items, how to decompose a complex item.
- **Past experience:** how long a special lasts, how long items take to sell out, who has good deals, what your competitors do.
- **Goals:** what the person wants, their tradeoff.
- **Observations:** what items are available, prices, number in stock,

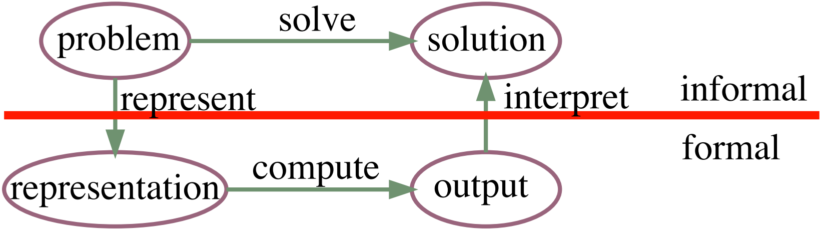
Intelligent Tutoring System

- Abilities: Present information, give tests
- Prior knowledge: subject material, primitive strategies
- Past experience: common errors, effects of teaching strategies
- Goals: the students should master subject material, gain social skills, study skills, inquisitiveness, interest
- Observations: test results, facial expressions, questions, what the student is concentrating on

Common Tasks of the Domains

- Modeling the environment Build models of the physical environment, patient, or information environment.
- Evidential reasoning or perception Given observations, determine what the world is like.
- Action Given a model of the world and a goal, determine what should be done.
- Learning from past experiences Learn about the specific case and the population of cases.

Representations



What do we want in a representation?

We want a representation to be

- rich enough to express the knowledge needed to solve the problem;
- as close to the problem as possible: compact, natural and maintainable;
- amenable to efficient computation
  - able to express features of the problem that can be exploited for computational gain
  - able to trade off accuracy and computation time and/or space
- able to be acquired from people, data and past experiences.

Defining a Solution

- Given an informal description of a problem, what is a solution?
- Typically much is left unspecified, but the unspecified parts can't be filled in arbitrarily.
- Much work in AI is motivated by common-sense reasoning. The computer needs to make common-sense conclusions about the unstated assumptions.

Quality of Solutions

- Does it matter if the answer is wrong or answers are missing?

Classes of solution:

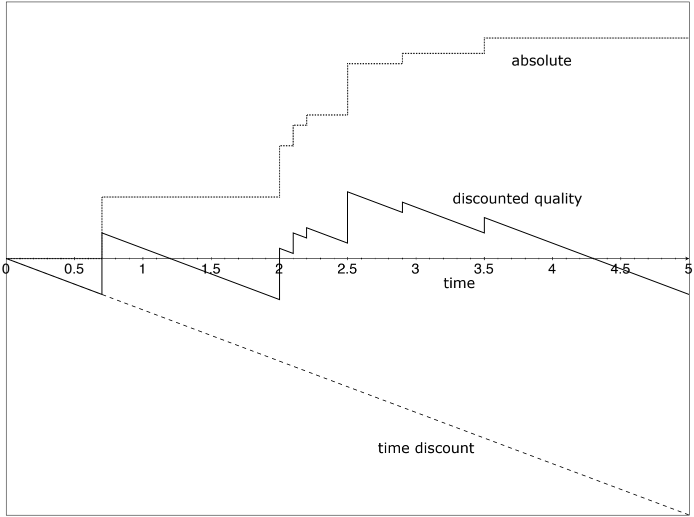
- An optimal solution is a best solution according some measure of solution quality.
- A satisficing solution is one that is good enough, according to some description of which solutions are adequate.
- An approximately optimal solution is one whose measure of quality is close to the best theoretically possible.
- A probable solution one that is likely to be a solution.

Decisions and Outcomes

- Good decisions can have bad outcomes. Bad decisions can have good outcomes.
- Information can be valuable because it leads to better decisions: value of information.
- We can often trade off computation time and solution quality.
  An anytime algorithm can provide a solution at any time; given more time it can produce better solutions.

An agent isn't just concerned about finding the right answer, but about acquiring the appropriate information, and computing it in a timely manner.

Solution quality and computation time



Reasoning and acting

Reasoning is the computation required to determine what an agent should do.

- Design time reasoning and computation is carried out by the designer of the agent.
- Offline computation is the computation done by the agent before it has to act.
  Background knowledge and data $\rightsquigarrow$ knowledge base.
- Online computation is the computation that's done by an agent between receiving information and acting.