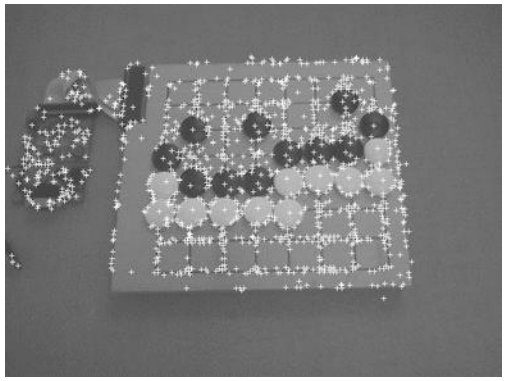
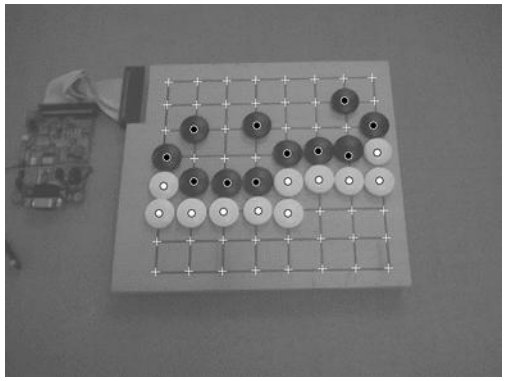
Artificial Intelligence Crash-Course





Alexander K. Seewald

SEEWALD SOLUTIONS

What is Artificial Intelligence?

Systems that think like humans	Systems that think rationally
"The exciting new effort to make computers think... machines with minds, in the full and literal sense" (Haugeland, 1985)	"The study of mental faculties through the use of computational models" (Charniak and McDermott, 1985)
"[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning..." (Bellman, 1978)	"The study of the computations that make it possible to perceive, reason and act" (Winston, 1992)
Systems that act like humans	**Systems that act rationally**
"The art of creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1990)	"A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes" (Schalkoff, 1990)
"The study of how to make computers do thinks at which, at the moment, people are better" (Rich and Knight, 1991)	"The branch of computer science that is concerned with the automation of intelligent behavior" (Luger and Stubblefield, 1993)

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Systems that think like humans

Cognitive Science

1960s Cognitive Revolution: information processing psychology replaced prevailing orthodoxy of behaviourism

Requires scientific theories of brain's internal activities

- Abstraction - level of Knowledge, Assemblies, Neurons...
- Validation - requires predicting and testing behavior of human subjects (top-down = Cognitive Science); and direct identification from neurological data (bottom-up = Cognitive Neuroscience)

Both approaches are distinct from AI but share direction. Much research on visual neuronal correlates of consciousness

Problem: (Prob.)Infeasible even for extremely small organisms

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Systems that think rationally

Laws of Thought

- Normative (or prescriptive) rather than descriptive.
- Aristotle: what are correct arguments / thought processes?
- Several Greek schools developed various forms of logic = notation and rules of derivation for thoughts; may or may not have proceeded to the idea of mechanization.

- Direct line via mathematics and philosophy to modern AI

Problems

- Not all intelligent behavior is related to logical deliberation
- The purpose of thinking = What thoughts should I have?
- Rational thinking is not possible without emotion

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Systems that act like humans

The Turing Test

Computing machinery and intelligence [Turing, 1950]

- Can machines think?⇒Can machines behave intelligently?
- Operational test for intelligent behavior = Imitation Game
- Pred. 30% chance for machine to fool lay person for 5mins
- Anticipated all major arguments against AI(!)

Suggested major components of AI: knowledge, reasoning, language understanding, learning

Problems

- Turing test is not reproducible and not constructive
- Chatbots based on (text-)mining terabyte of chat room logs are often judged intelligent by non-experts

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Systems that act rationally

Doing the right thing

- Rational behaviour: doing the right thing
- The right thing: which is expected to maximize goal achievement given the available information
- Doesn't necessarily involve thinking, but thinking should be in the service of rational action.

Aristotle (Nicomachean Ethics)

Every art and every inquiry, and similarly every action and pursuit, is thought to aim at some good.

Problem

- Definition of good - must be efficent and fast to check for the agent and still be compatible with complex human definitions

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

AI Prehistory

Philosophy	logic, methods of reasoning mind as physical system foundations of learning, language, rationality
Mathematics	formal representation and proof algorithms, computation, (un)decidability, (in)tractability, probability
Psychology	adaptation, phenomena of perception and motor control, experimental techniques
Economics	formal theory of rational decisions
Linguistics	knowledge representation, grammar
Neuroscience	plastic physical substract for mental activity
Control Theory	homeostatic systems, stability simple optimal agent designs

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

AI History

1943	McCulloch & Pitts: Boolean circuit model of brain
1950	Turing's *Computing Machinery and Intelligence*
1952-69	Look, Ma, no hands! - Phase
1950s	Early AI programs: Samuel's checkers, Newell & Simon's Logic Theorist; Winograd's Blocks World
1956	Dartmouth meeting: Artificial Intelligence adopted
1965	Robinsons complete logical reasoning algorithm
1966-74	AI discovers computational complexity
1969-79	Early development of knowledge-based systems
1980-88	Expert systems industry booms
1988-93	Expert systems industry busts: "AI Winter"
1988-	Resurgence of probability; increase in technical depth "Nouvelle AI": ALife, Genetic Algorithms, soft computing
1995-	Agents metaphor; real-world applications

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Recent AI Successes

2005 Stanford's Stanley wins DARPA Grand Challenge, driving autonomously 131 miles through the desert (but blind - no camera!)

2010 Kinect 360° motion sensor uses major AI research area *Machine Learning* to recognize body parts from depth information w/o calibr.

2011 IBM's Watson beats the two greatest Jeopardy! champions

2012 Geoffrey Hinton: Speech recognition / Deep Learning starts AI hype

2015 Tesla announces software update to enable self-driving on freeways
Self-driving cars now mainstream (but not all problems resolved...)

2016 Google DeepMind AlphaGo beats 9dan Go champion Lee Sedol 4:1

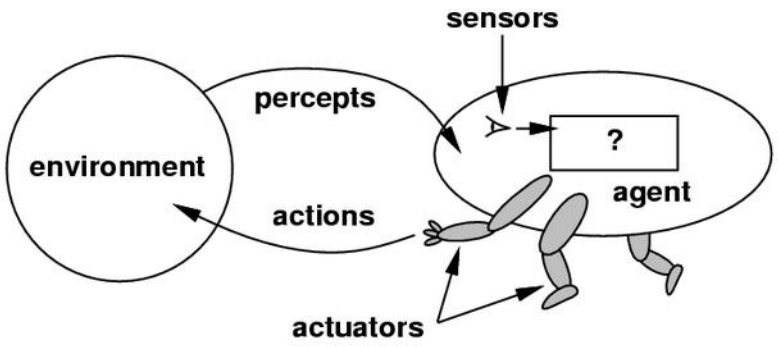
2017 AlphaZero: Chess&Shogi at champion level in days
5^{th} Schur no. S(5) = 161 (proof: 2 petabytes, 3 days, 14 CPU years)

2019 AlphaStar: Realtime strategy game Starcraft II - 10:1 vs. human opp.

Successes are due to increases in computing power, improvement in learning algorithms, greater emphasis on solving sub-problems, and collaboration with related fields. Many tasks are still *AI-complete*, no human-level intelligence!

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Agents and environments



An agent is everything that perceives and acts.

The whole field of AI can be viewed as being concerned with design of intelligent embodied agents.

Agents include humans, robots, softbots, vacuums cleaners...

The agent function maps from percept histories to actions:

$$f: P^* \rightarrow A$$

For any given class of environments and tasks, we seek the agent with the best performance. Computational limitations make perfect rationality unachievable.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Types of agents

Four basic agent types in order of increasing generality:

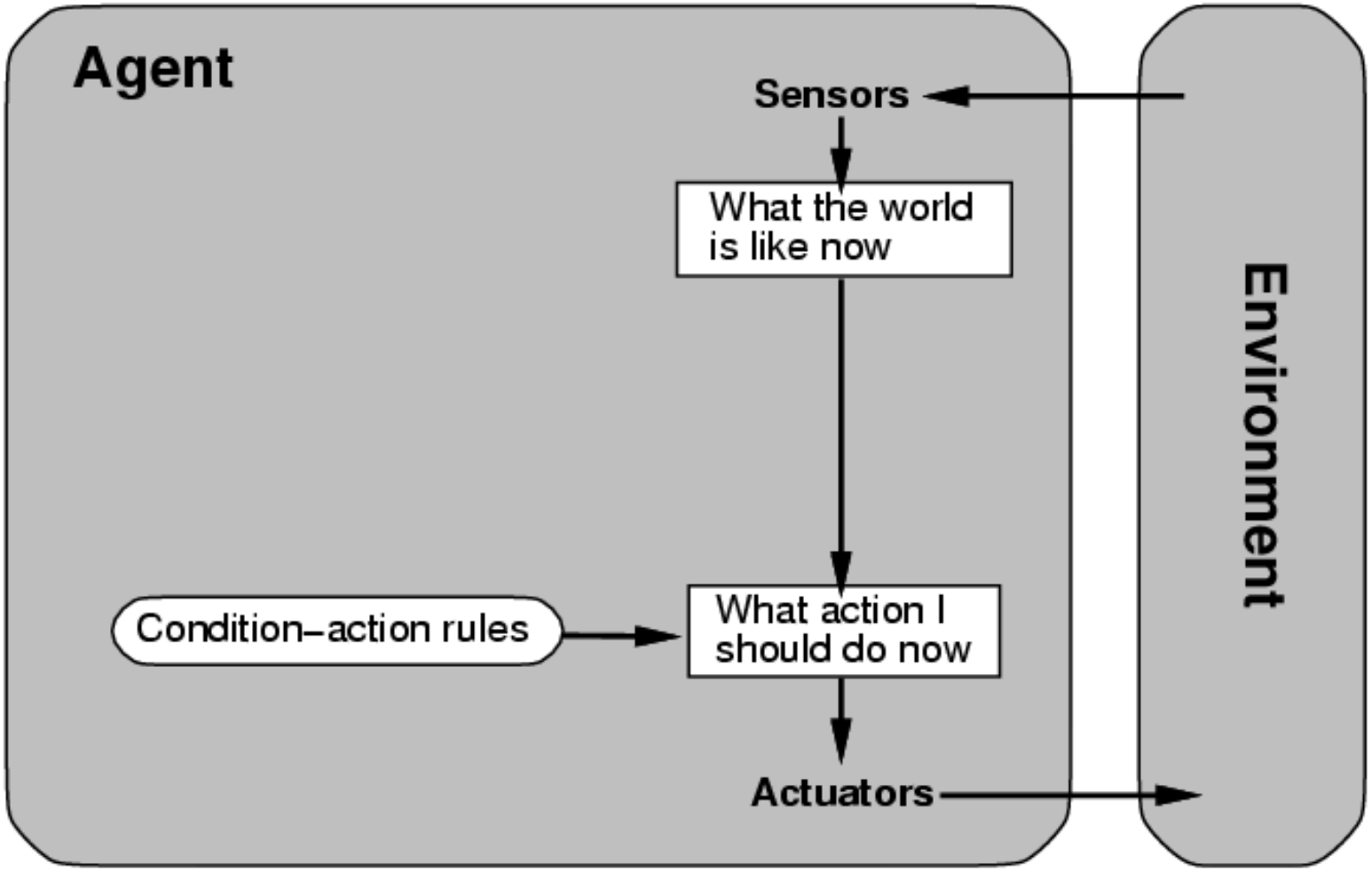
- Simple reflex agent
- Reflex agent with state
- Goal-based agent
- Utility-based agent

All these can be turned into learning agents, where some aspects of the agent can be changed by experience.

Learning is the central issue for intelligent agents. The research fields of Machine Learning and Data Mining have investigated simpler learning model for decades. While a general learning agent is still decades away, ML & DM are well on the way towards a mature field.

alex@seewald.at
http://www.seewald.at

Simple reflex agent



SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Example: Vacuum cleaner agent

Percepts: clean/dirty, wall, stairs

Actions: move, rotate, clean

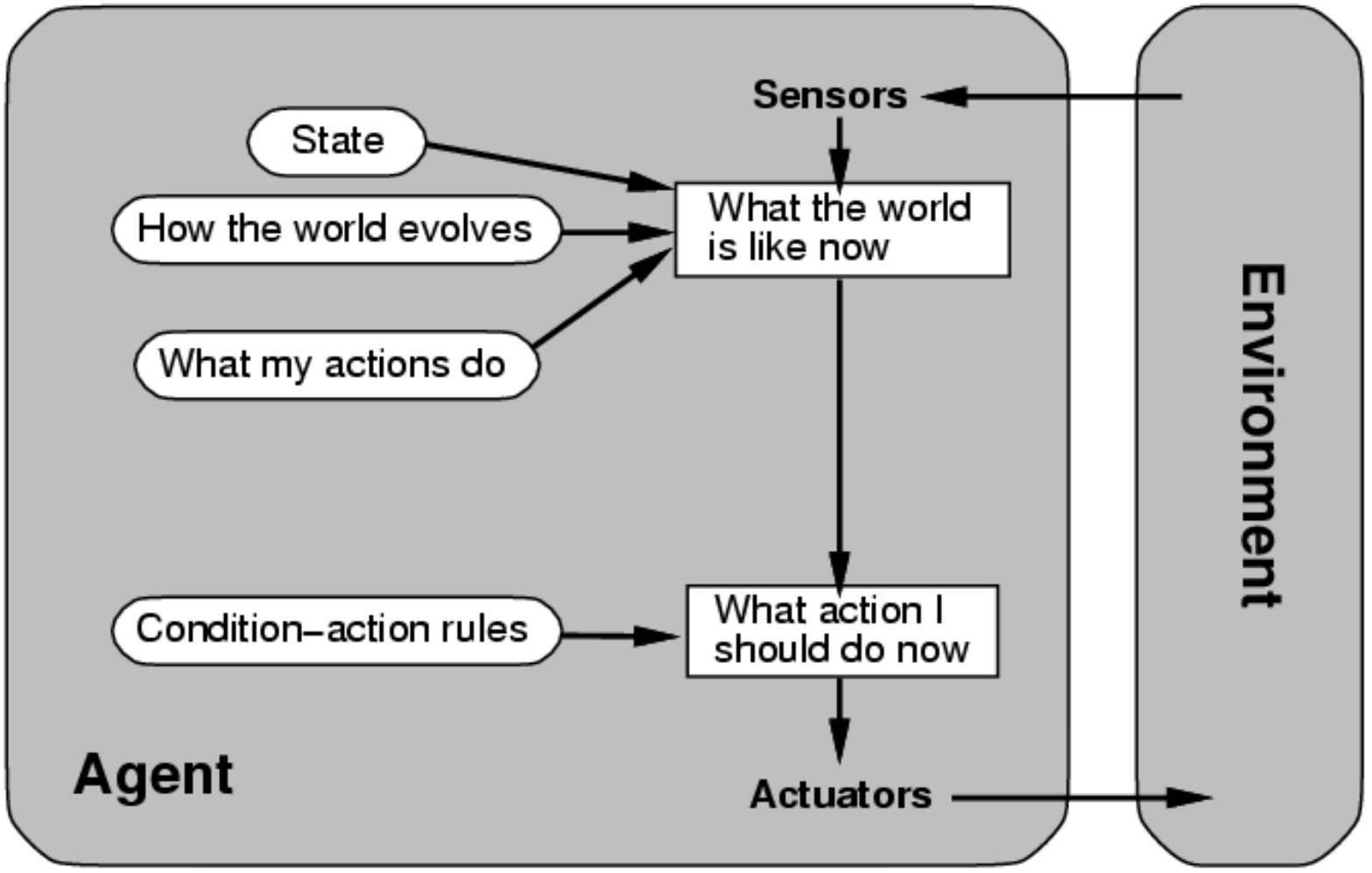
Goals: maximize amount of dirt collected / cleanliness

Environment: single-level household

SEEWALD SOLUTIONS

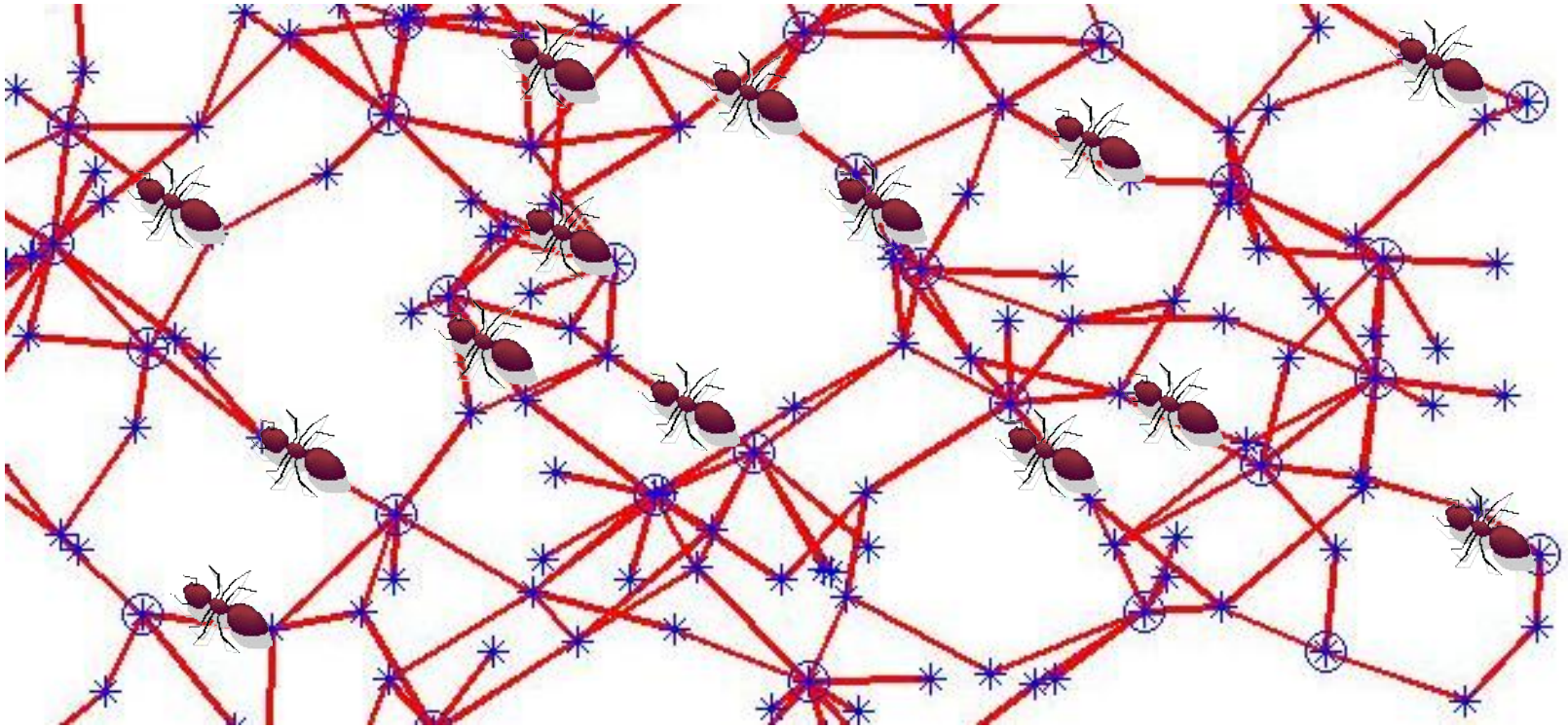
alex@seewald.at
http://www.seewald.at

Reflex agent with state



alex@seewald.at
http://www.seewald.at

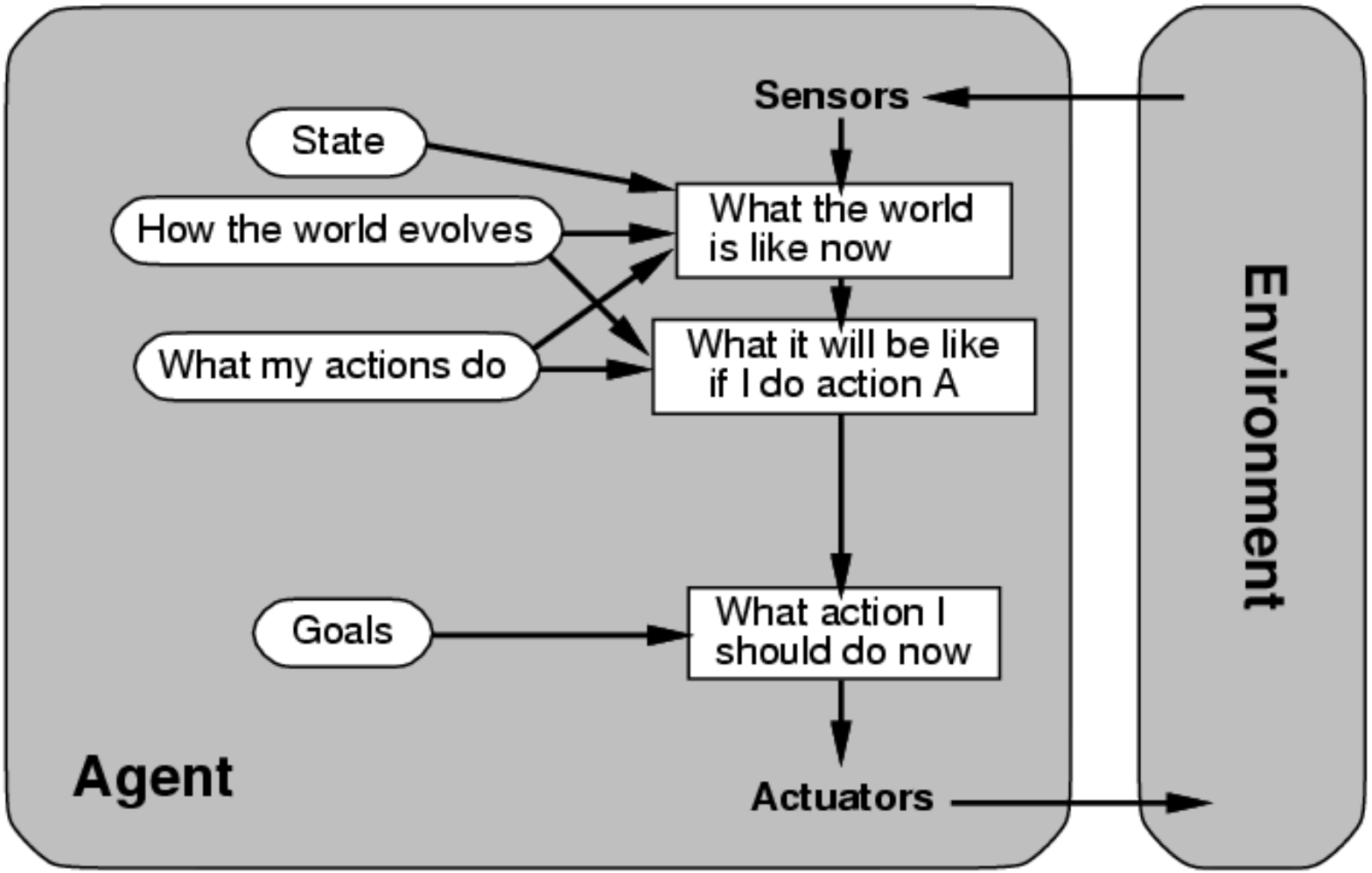
Example: Ant-based routing



[Di Caro & Dorigo, 1998] have shown that ant-based routing outperforms other common routing methods. State is the history of visited nodes; similar to pheromone tracks in real ants.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

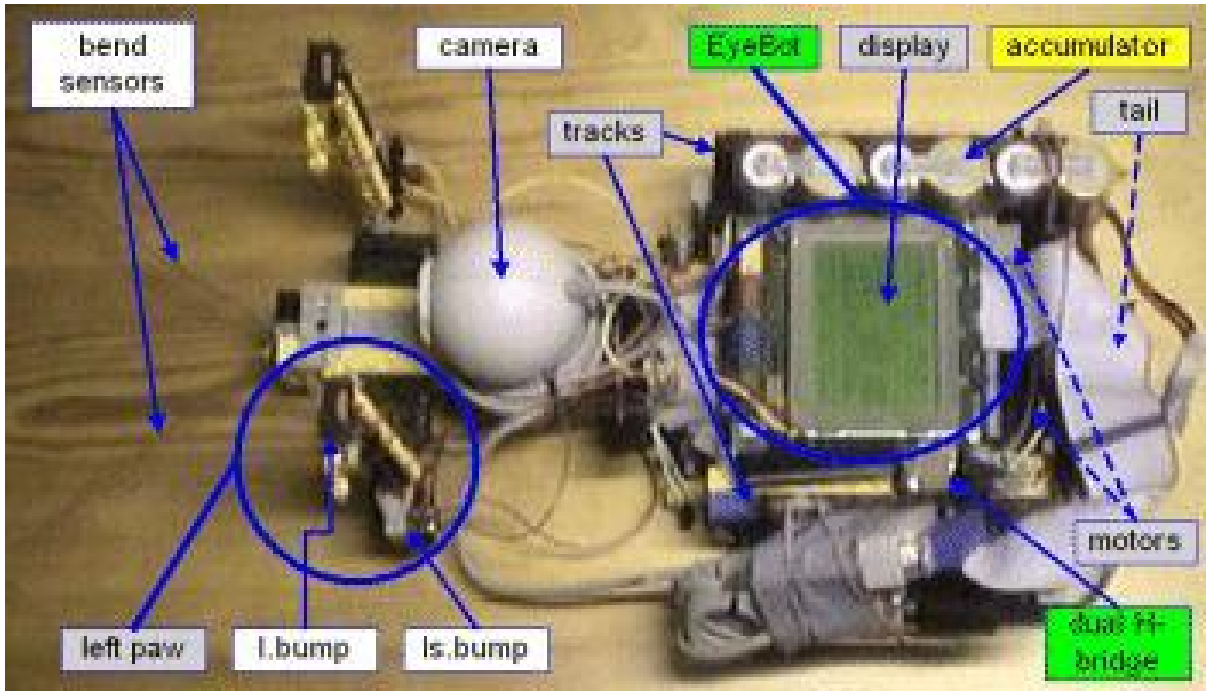
Goal-based agent



SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

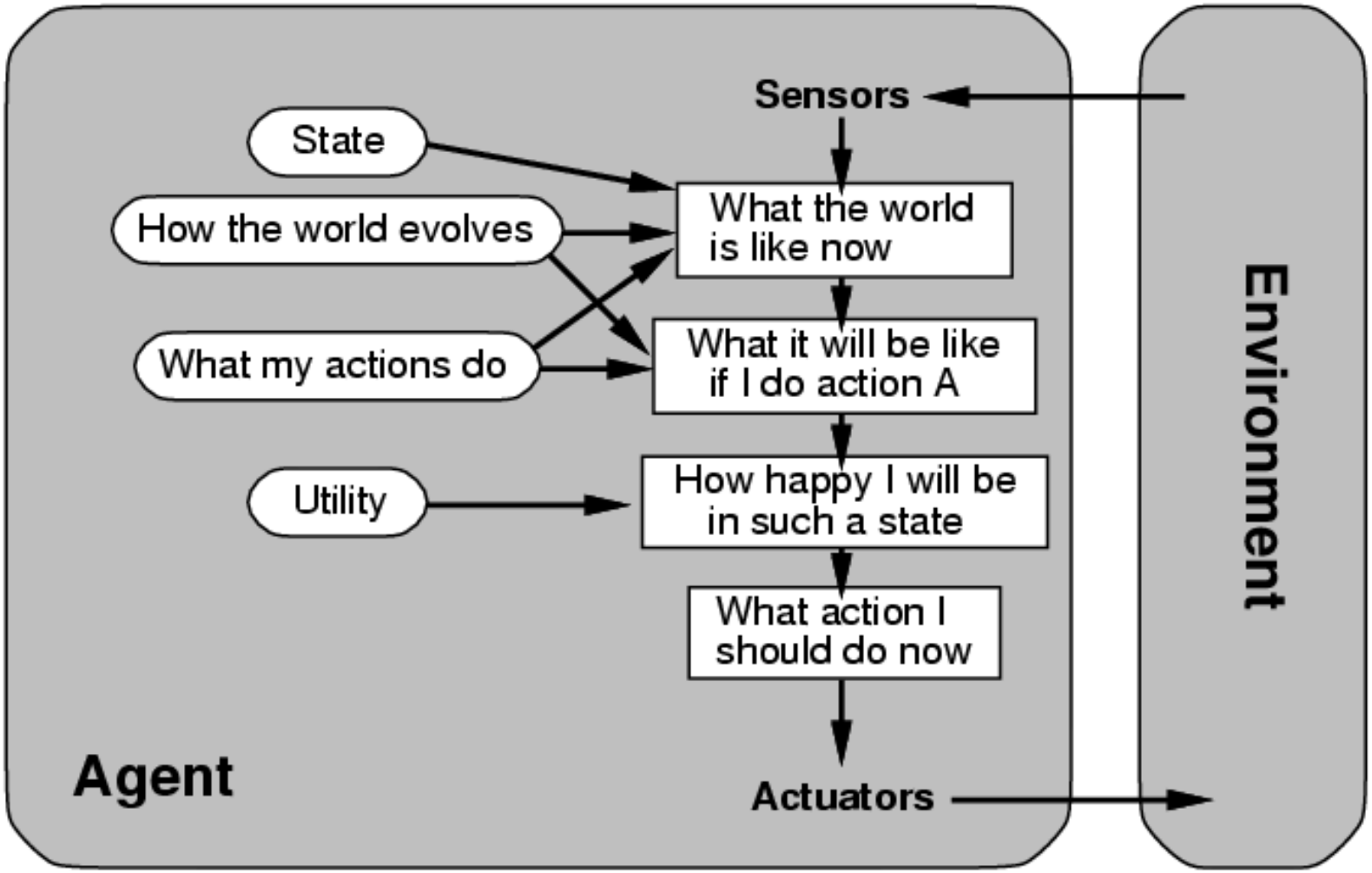
Example: RoboCat



RoboCat (Seewald, 1999; Diploma thesis) is an example for a robot with is controlled by emotion and motivation, and shows realistic non-deterministic behaviour. **Video**

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Utility-based agent



SEEWALD SOLUTIONS

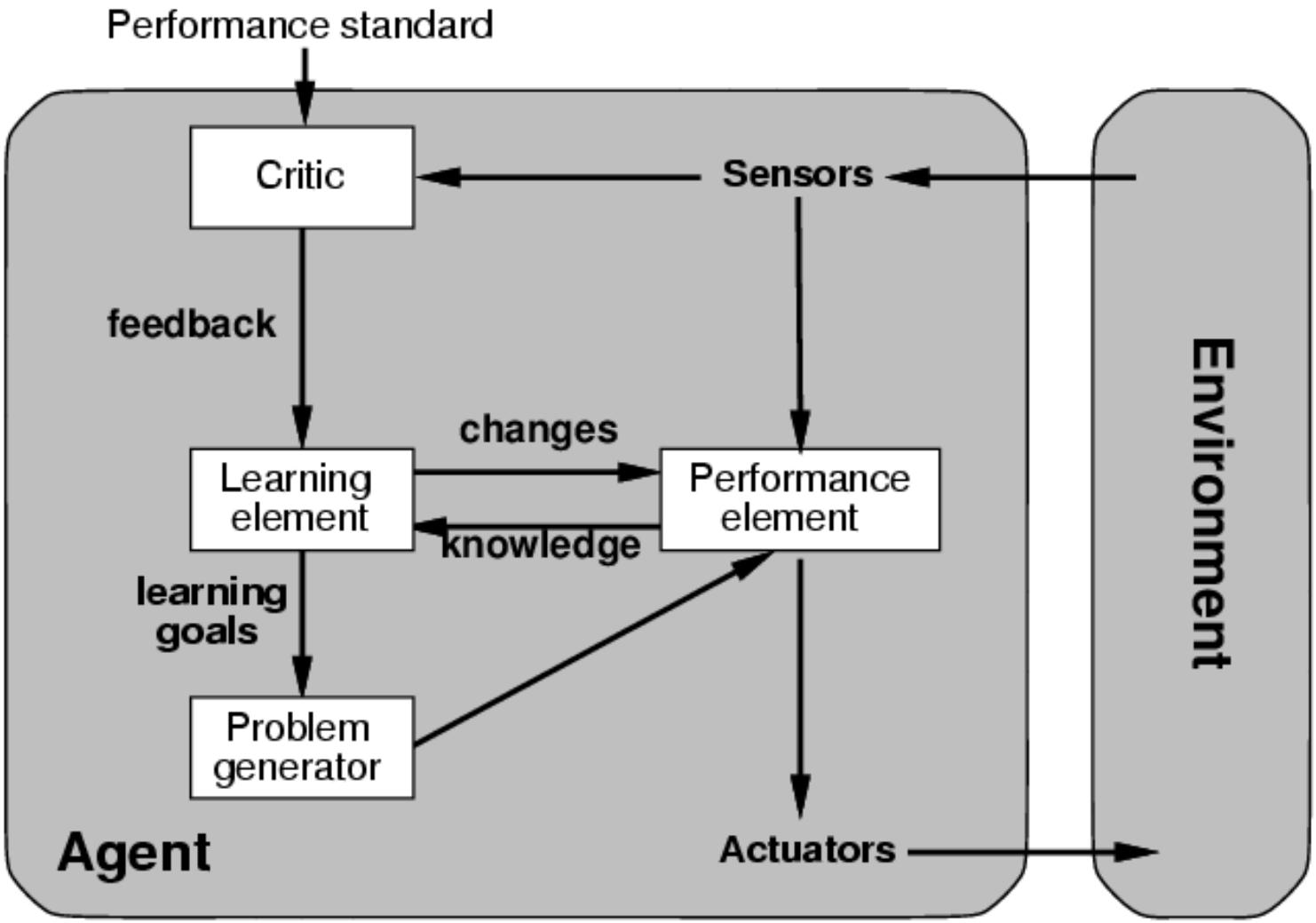
alex@seewald.at
http://www.seewald.at

Example: Invisible Person

The Invisible Person project (1999-2005) with the Technical Museum in Vienna was concerned with the creation of an engaging playful agent. **Video**

alex@seewald.at
http://www.seewald.at

Simple Learning Agent (reflex-based)



SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Example: Stanley

Autonomous robot vehicle which won the DARPA Challenge 2005. Built at Stanford University in about 15 months by a team of around 35 people. Uses Machine-Learned Laser Perception and Speed Strategy.

Presenting Stanley

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Stanley

Sensors

- Sight: Five laser rangefinders; monocular video camera; radar for long range sight
- Position: GPS sensor with 20cm resolution for pose estimation; measurements of wheel speed for pose estimation
- Balance: a 6 DOF inertial measurement unit; GPS compass generates 2 DOF balance information from two separate GPS antennas

Brains

- Six Pentium M motherboards in a rugged rack mount unit
- Battery-backed, electronically-controlled power system
- Custom software modules for: Planning and Optimization; Control; LIDAR - Light Detection and Ranging; Computer Vision; Inertial Navigation; Reliability
- Data sampling from instruments at rates varying from 10Hz to 100Hz

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Vision System

Adaptive Vision

- Use laser range finder to locate a smooth patch of ground ahead
- Sample color and texture of this patch from monocular video image
- Scan for same color and texture in the whole image

⇒ *Road Segmentation*

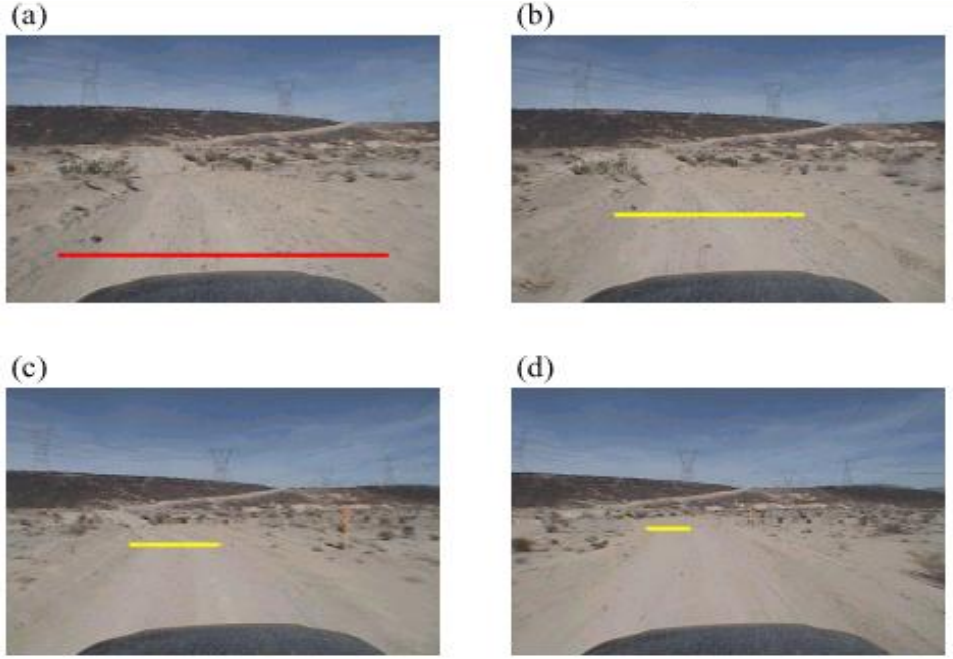
SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Alternative Vision System

"Adaptive Road Following using Self-Supervised Learning and Reverse Optical Flow" (Lieb, Lookingbill & Thrun, 2005) [not used in DARPA Challenge]

- Assume: region directly in front of the vehicle is drivable road.
- Sample region at various distances from past images, using reverse optical flow to determine its previous position.
- Match each sampled region at appropriate vertical pos.in the current image
- Integrate via Dyn.Prog.

⇒ *Road Segmentation*



alex@seewald.at
http://www.seewald.at

Tesla Model S

Auto-Pilot by Software Update in Mid-Oct.2015 based on existing hardware

- Front-facing camera on top of windshield
- Forward-facing radar in lower grill
- Ultrasonic sensors in front and rear bumpers for 360° "view"
- Trained via *end-to-end deep learning*

***Not* sufficient for completely autonomous driving!**

- Only works on freeways with very good lane markings or with much car traffic
- Snow, strong rain, darkness is still an issue when recognizing lane markings
  Lane markings not always present, hard to recognize or appear multiple times
- Can read *some* traffic signs (e.g. speed limit)
- Not aware of two-way traffic (**Video**)

SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Example: AI in Finance

At present

- Estimation of creditworthiness (logistic regression / everyone)
- High-speed trading systems capturing arbitrage (Jane Street Capital)
- 24-7 customer service via chatbots (Personetics, Gridspace; Bank of America: Erica)
- Analyzing commercial loan contracts (JPMorgan COIN)

In the future

- Fully automated 24-7 auditing - flagging suspicious transactions instantly as they occur
- Better estimation of creditworthiness using unstructured data (tweets, facebook posts, snapchat images, ...)

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

How can we build such agents?

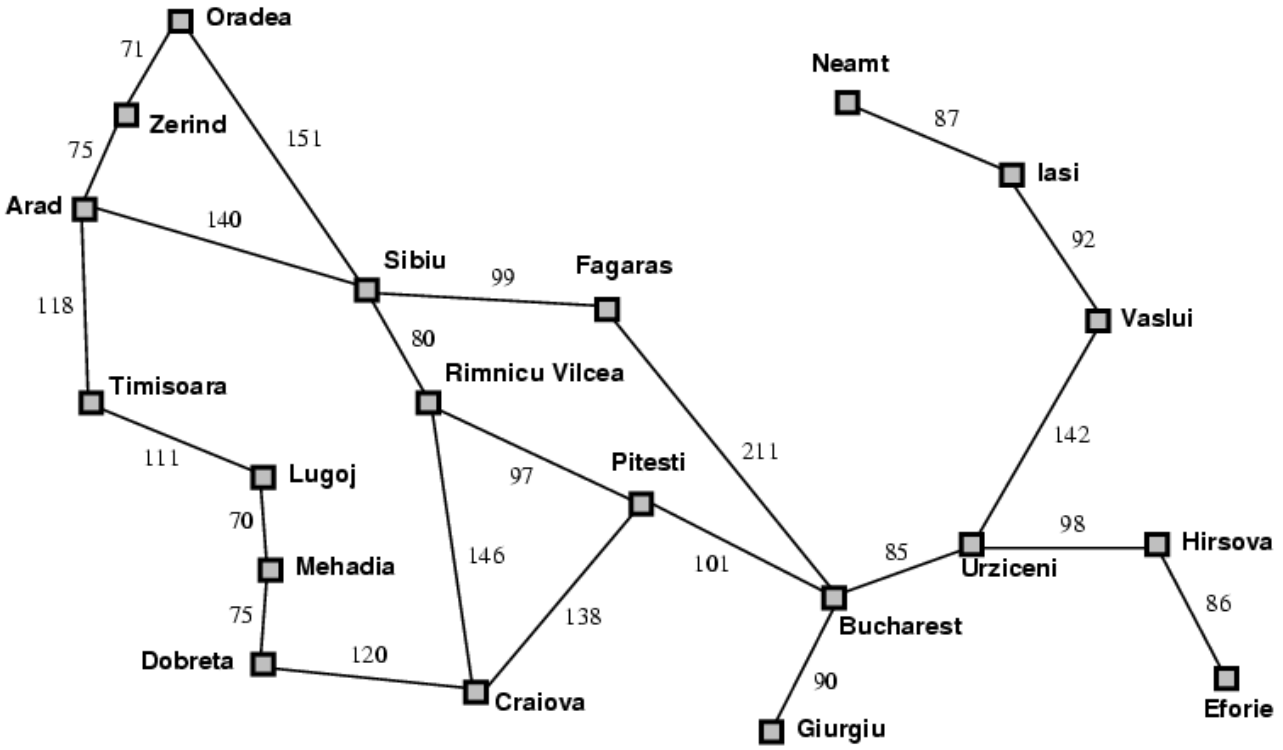
- Search / Problem Solving
- Knowledge and Reasoning
- Acting under Uncertainty
- Decision Theory
- Communication / NLP

Learning

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Search / Problem Solving

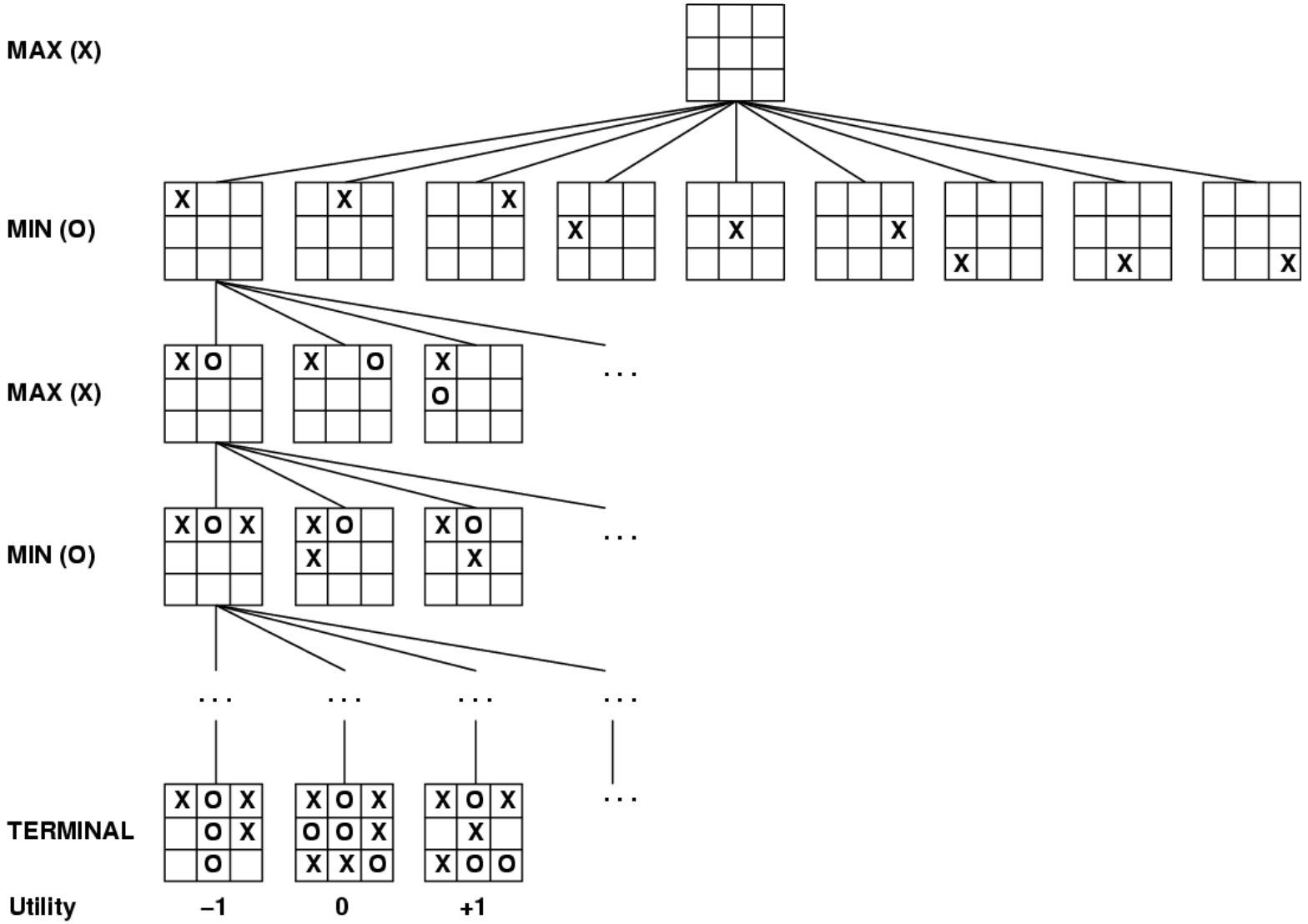
Search is a central theme in AI. The fastest path through a city; VLSI layout; the correct interpretation of a given sentence; and even general learning - all these can be formulated as search problems.



SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Example: Game as Search



alex@seewald.at
http://www.seewald.at

Knowledge and Reasoning

Intelligent agents need **knowledge** about the world in order to reach good decisions. Humans use huge amounts of implicit **common-sense knowledge** to solve even tiny tasks. All attempts to model this knowledge have failed.

Constructing **knowledge-based systems** has advantages over programming, but is not feasible for all problems. Modeling relevant knowledge for a task may be infeasible.

State-of-the-Art are **embedded AI** systems, where AI is used complementary to other programming techniques.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Example: RoboSail Systems

- Autopilot for one-person sailing
- Race-proven with various state-of-the-art AI and ML components.
- Human jargon like *gust*, *close-hauled*, *luff* as background knowledge!

SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Planning

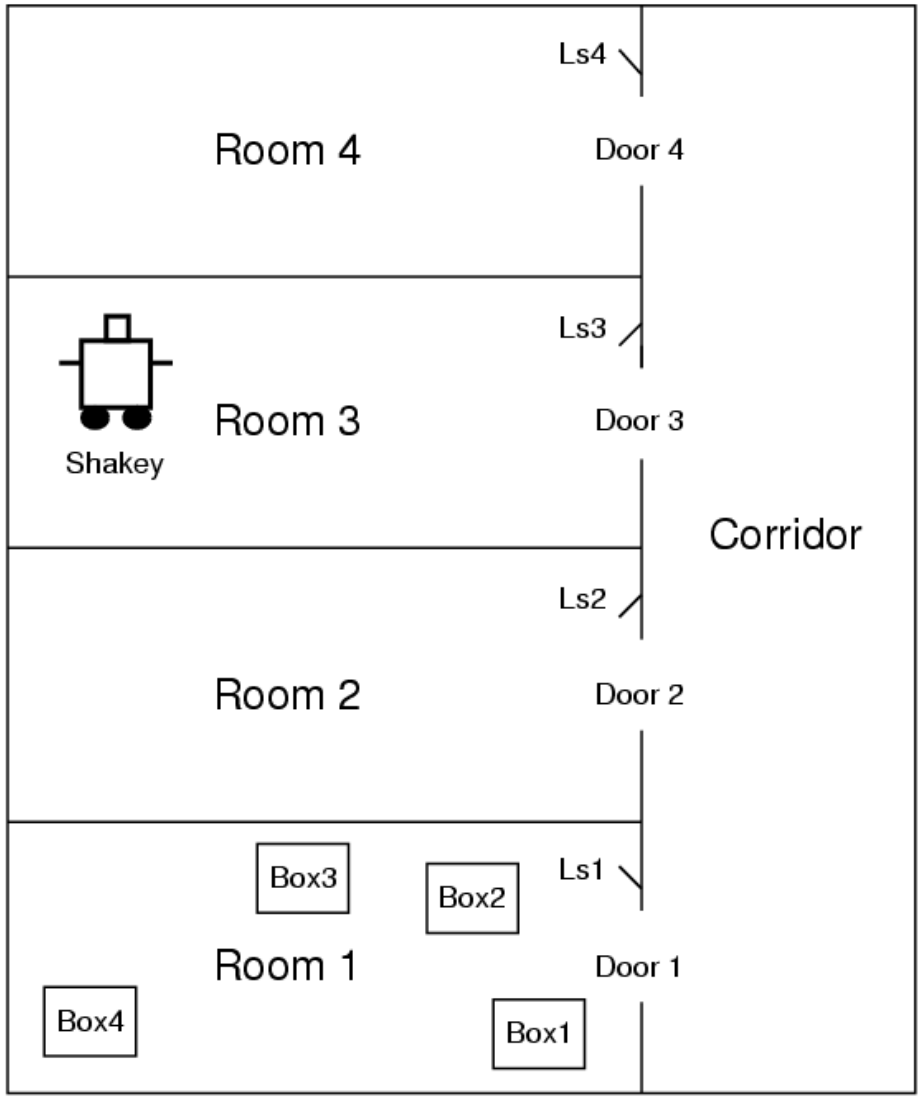
Planning agents *look ahead* to come up with actions that will contribute to goal achievement. They differ from problem-solving agents in their use of more flexible representations of state, actions, goals, and plans.

Planning systems can be seen as efficient special-purpose reasoning systems designed to reason about actions; or as efficient search algorithms for the space of possible plans.

Automatic planners and schedulers have proven capable of handling complex domains such as spacecraft missions and manufacturing

alex@seewald.at
http://www.seewald.at

Example: Shakey



SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

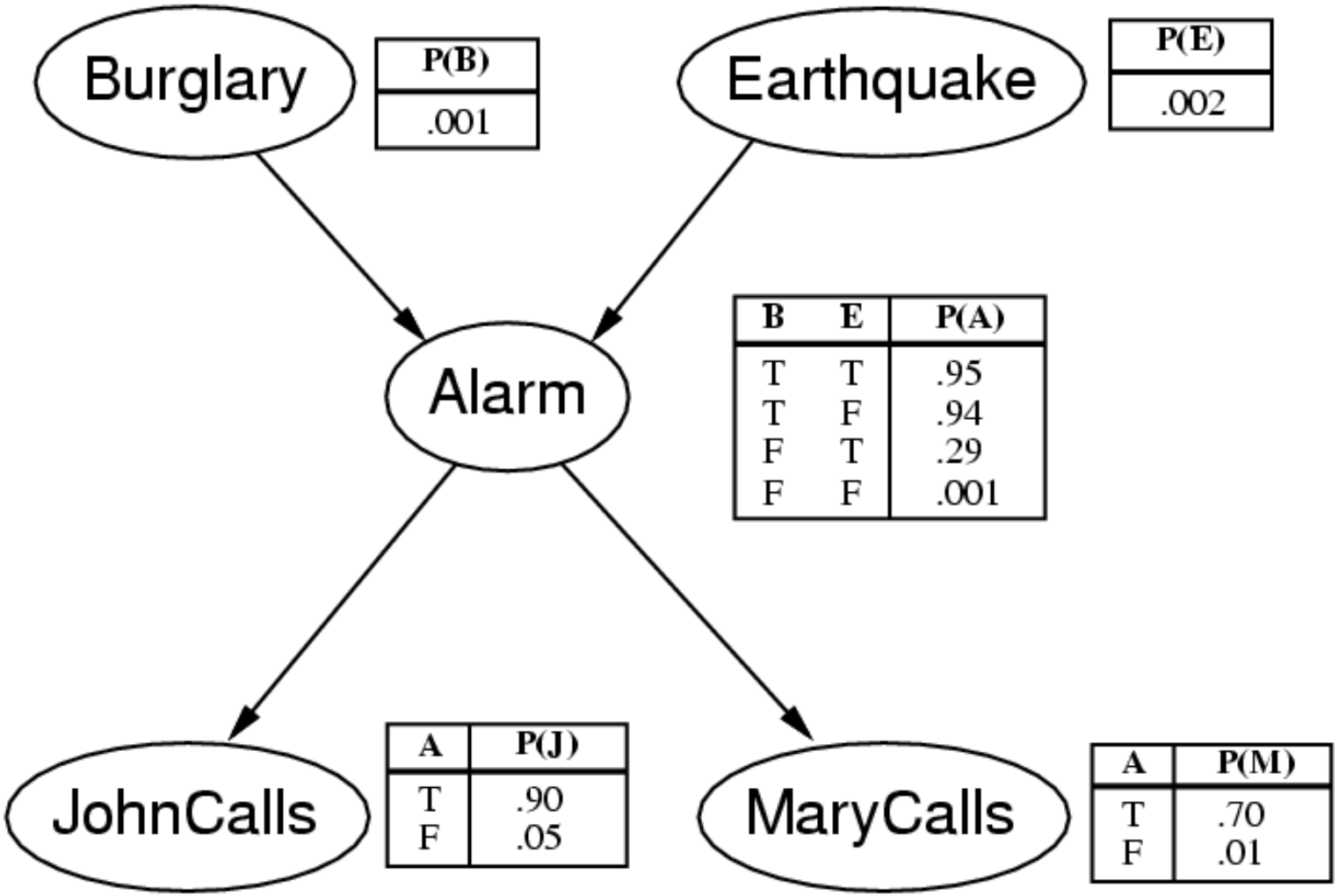
Acting under Uncertainty

Uncertainty is inescapable in complex, dynamic or inaccessible worlds; and means that many simplifications that are possible with deductive inference are no longer valid. **Probability theory** provides a way of summarizing the uncertainty that comes from laziness and ignorance.

Belief networks are a natural way to represent conditional independence information. The links between nodes represent the qualitative aspects of the domain, and the conditional probability tables represent the quantitative aspects.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Example: Burglar alarm



SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Decision Theory

Simple decision problems can be solved by **decision theory**, which relates what an agent wants (**utility theory**) to what an agent should believe on the basis of evidence (**probability theory**) Utility theory associates a utility value to each state of the agent.

We can use decision theory to build a system that make decisions by considering all possible actions and choosing the one that leads to the best expected outcome. Such a system is known as a **rational agent**.

Decision theory is **normative** - it describes rational behaviour. It is definitely not **descriptive** - people systematically violate the axioms of utility theory.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Question to the Audience

What would you prefer?

A) 80% chance of winning €4000

B) 100% chance of winning €3000

[Allais, 1953] found that people strongly prefer B)

C) 20% chance of winning €4000

D) 25% chance of winning €3000

[Allais, 1953] found that people strongly prefer C)

Inconsistent human utility theory on monetary value!

$0.8U(€4000)<U(€3000)$ and $0.25U(€3000)<0.2U(€4000)$ cannot both be satisfied.

SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Communication

Agents need to communicate to each other and to the users. Communication between learning agents is an active research area which sheds light on the development of language in humans.

Natural language processing techniques make it practical to develop programs that make queries to a database, extract information from texts, translate languages, or recognize spoken words.

In all these areas, there exist programs that are useful, but there are no programs that do a throrough job in an open-ended domain.

alex@seewald.at
http://www.seewald.at

Example: Alexa

Amazon's Echo NLP user interface

- Personal assistant with speech-based interface (speech data stored and analyzed in the cloud)
- 25 Million units actively used (only US)
- 15,000 Skills (voice-based apps)
- Skills available by Ally Bank, American Express and several other banks: get bank balance, transfer money, get rates, hear recent transactions....
- No user authentication - anyone can give all commands (if pin is not configured)

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Example: Mastercard Decision Intelligence

- Generates one predictive real-time fraud score for every credit card transaction
- Distinguishes normal and abnormal spending behaviours from historical data
- Data points used: IP-address of purchasing device, device identification, phone number, email adress
- Modifies normal fraud model towards higher transaction approvability

Goal: To reduce number of transactions falsely declined without increasing overall fraud incidence

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Agents as programming metaphor

- Procedural (classic) programming
- Declarative programming
- Object-oriented programming
- Constraint logic programming
- Event-oriented programming
- Knowledge-based software engineering
- Agent-based software engineering

...

Each of these gives an unique viewpoint on programming; makes solving some problems easier and others harder. *But* you still need a programmer!

For learning systems, you don't need a programmer. Most of the work is done by learning systems.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

MACHINE LEARNING

„The field of machine learning is concerned with the questions of how to construct computer programs that automatically improve with experience.“ (Tom M. Mitchell, 1997)

DATA MINING

„Data Mining is the non-trivial process of identifiying valid, novel, potentially useful, and ultimately understandable patterns in data.“ (Fayyad Piatetsky-Shapiro & Smyth, 1996)

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Learning (1)

A large variety of learning systems is available which can learn...

- A state evaluation function to play checkers
- A belief network to model sleep stages
- A function to predict insurance risks
- Logic programs to determine cancerogenity
- Association rules in supermarket basket analysis
- Time-dependent models of speech
- Shapes of biological objects (e.g. erythrocytes)
- The color and texture of a walkway in a park

...

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Learning (2)

Learning a function from examples of its inputs and outputs is called **inductive learning**. Learning in the inductive setting is supervised and needs a set of training inputs and outputs.

Unsupervised learning uses the structure of training data to infer hidden relationships, which are harder to validate.

Inductive logic programming can learn relational knowledge, as used in knowledge-based systems. This kind of learning is generally very hard for larger problems.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Learning (3)

Learning in intelligent agents is essential for dealing with unknown environments; and for building agents without prohibitive amount of work. All learning suffers from the **credit assignment** problem = which steps are responsible for a good or bad outcome?

Reinforcement learning is an active research topic, and computationally very expensive. Temporal difference learning and Q-Learning are common learning algorithms.

Genetic algorithms achieve reinforcement by increasing the proportion of successful functions. They achieve generalization by mutating and cross-breeding programs.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Learning

Learning a function from examples of its inputs and outputs is called **inductive learning**. Learning in the inductive setting is supervised and needs a set of training inputs and outputs.

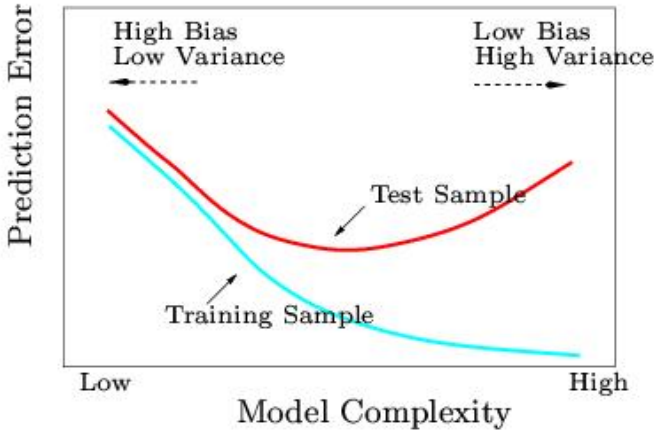
Unsupervised learning uses the structure of training data to infer hidden relationships, which are harder to validate.

Learning in intelligent agents is essential for dealing with unknown environments; and for building agents without prohibitive amount of work. All learning suffers from the **credit assignment** problem = which steps are responsible for a good or bad outcome?

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Bias

"Bias refers to any criterion for choosing one generalization over another other than strict consistency with the observed training instances" (Mitchell, 1980)



Each learning algorithm is biased twofold:

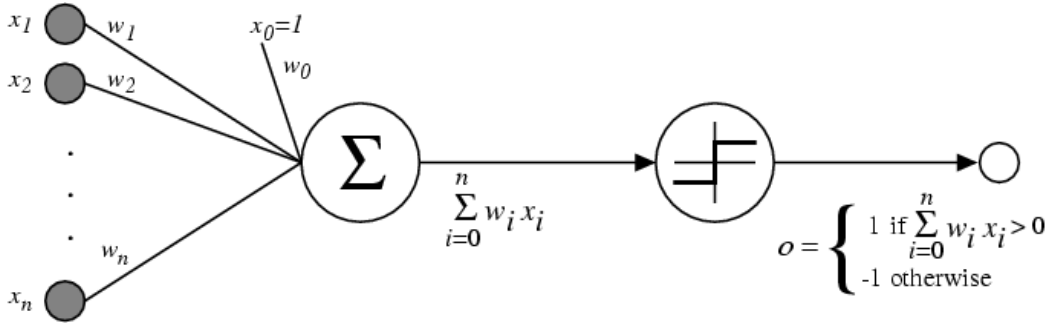
- **language bias** = restricts possible concepts to be learned
- **search bias** = prefers certain models over others

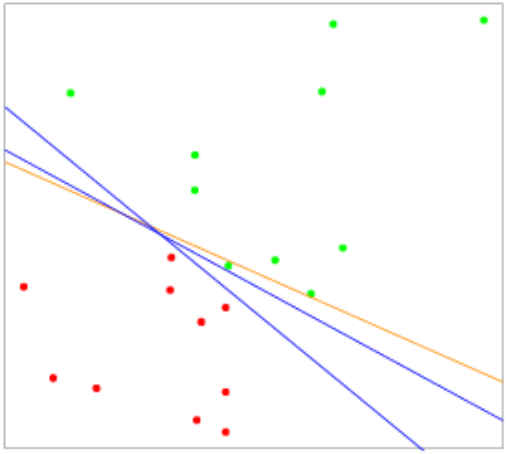
Overfitting occurs when the structure of training data is learned too well; and the generalization performance on unseen data suffers.

Bias is essential to learning!

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Deep Learning (1)



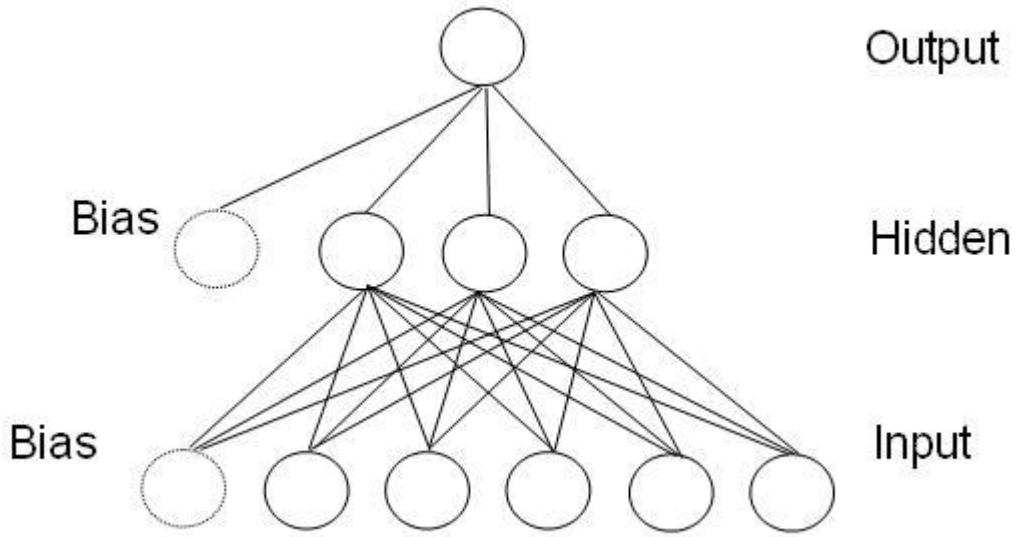


Perceptron (linear binary threshold unit)

- Computes a linear function of **x** (assume adding an $x_0=1$ to **x**, so that constant term w_0 can be handled). $f(\mathbf{x}) = \text{sign}(\mathbf{x}^T.\mathbf{w})$. **w** is initialized randomly.
- *Perceptron training rule*: $\mathbf{w} \leftarrow \mathbf{w}+\eta(y-f(\mathbf{x})).\mathbf{x}^T$, where y is the true output value from training data (±1), and η is the learning rate.
- Concept boundary is a hyperplane which separates classes +1 & -1.
- Very similar to linear regression but does not converge in all cases

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Deep Learning (2)

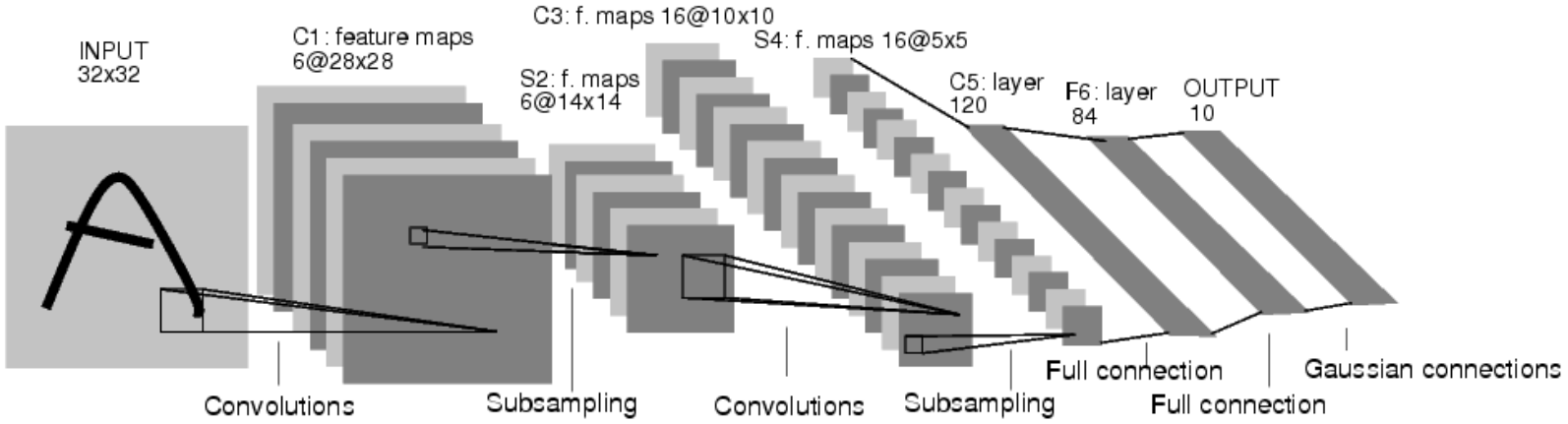


Multi-Layer Perceptron

- Kombination vieler Perceptrons
- Lernen (=Gewichte optimieren) mit Backpropagation
- Explizite Bias-Knoten mit konstantem Outputwert
- Kann auch nichtlineare Konzepte (wie XOR) lernen

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Deep Learning (3)



Convolutional Neural Network

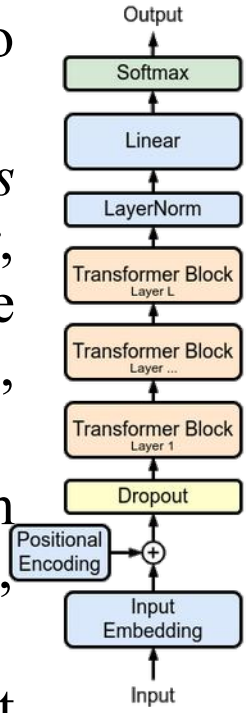
- Subsampling: Mittelwert von benachbarten Pixeln
- Convolution: Inputs = benachbarte Pixels, Gewichte für alle Knoten gleich (weight sharing)
- Lernen mit adaptierter Backpropagation
- Ursprünglich sehr instabiler Lernalgorithmus - in den letzten 5 Jahren deutlich verbessert und beschleunigt (jetzt: Deep Learning)
- Performance hängt stark von Netzwerkstruktur ab

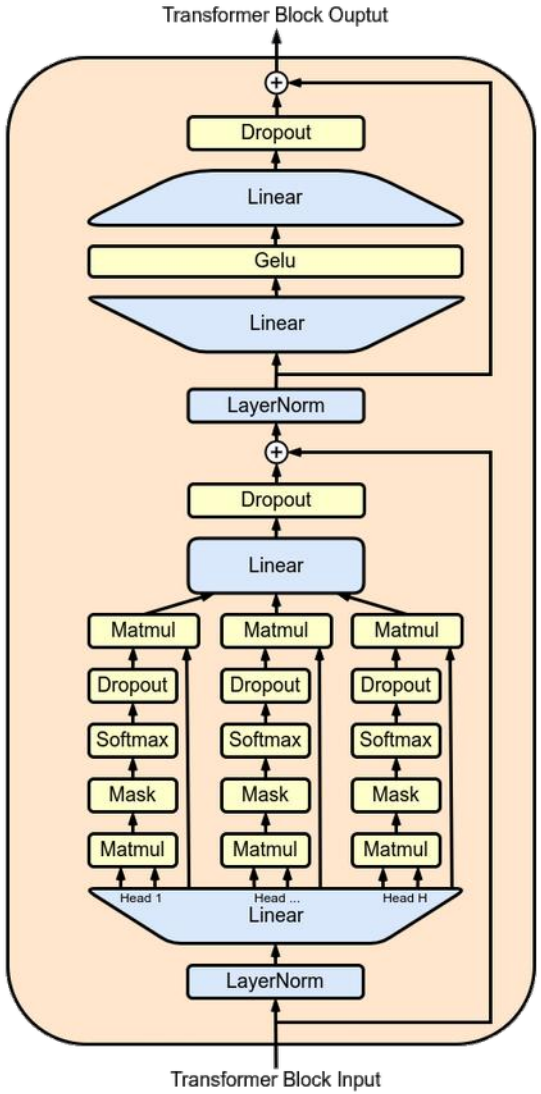
SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Deep Learning (4)

Generative Models

- Maps non-structured data to non-structured data (e.g. text to text, text to image, image to text, text to sound, ...)
- Learn to generate *completions* of text (question → answer, finish sentences, give summaries etc.. e.g. GPT-4, BERT, LaMDA,...)
- Learn to generate images from caption text (e.g. Midjourney, DALL-E, Stable Diffusion)
- Very trick to train, needs vast amounts of training data plus extensive human feedback





SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Deep Learning (5)

Major improvement on state-of-the-art as of 2012!

- Can learn from **any** kind of unstructured data **directly** (images, video, audio, writing, time series, text, source code, play actions, ..)
- Does its own feature construction and selection
- Models for supervised learning are usually learned extremely well ~ super-human performance with enough training data
- Even very small models show surprisingly complex behaviours
- Models can be applied very efficiently on less powerful platforms
- Well-supported code-base, also to be part of next C++ standard

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Deep Learning - Caveats (6)

Solvable Caveats

- Needs *millions to billions* of training samples which is impossible to get for most real-world applications - must *extend* samples using ingenious methods (e.g. handwritten digit recognition: elastic distortion of real digits)
- Very large number of network architectures exist! Only for very few tasks are well-performing network architectures known (U-net)
- Training the network takes a long time. GPU-optimized versions or large computer clusters are highly recommended, especially when determining network architecture. Testing is normally fast and can sometimes run on embedded hardware. Pretrained models for related fields can reduce training time significantly.

- Simple backpropagation is still the best learning method if you can spare the much longer training time.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Deep Learning - Caveats (7)

(Probably) Unsolvable Caveats

- Winograd Schemas: simple problems in pronoun disambiguation which need commonsense knowledge to solve (~ AI-complete)

The city councilmen refused the demonstrators a permit because they [feared/advocated] violence.

The cat was lying by the mouse hole waiting for the mouse, but it was too [cautious/impatient].

Bob collapsed on the sidewalk. Soon he saw Carl coming to help. He was very [ill/concerned].

The dog chased the cat, which ran up a tree. It waited at the [top/bottom].

More schemas here:

https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WSCollection.html

The generation of such schemas is an open research problem.

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at

Deep Learning - Caveats (8)

Known (probably unsolvable) Caveats

- ColorFool (Semantic Adverserial Colorization)

Simple changes in back-ground color yield strange changes in image class.

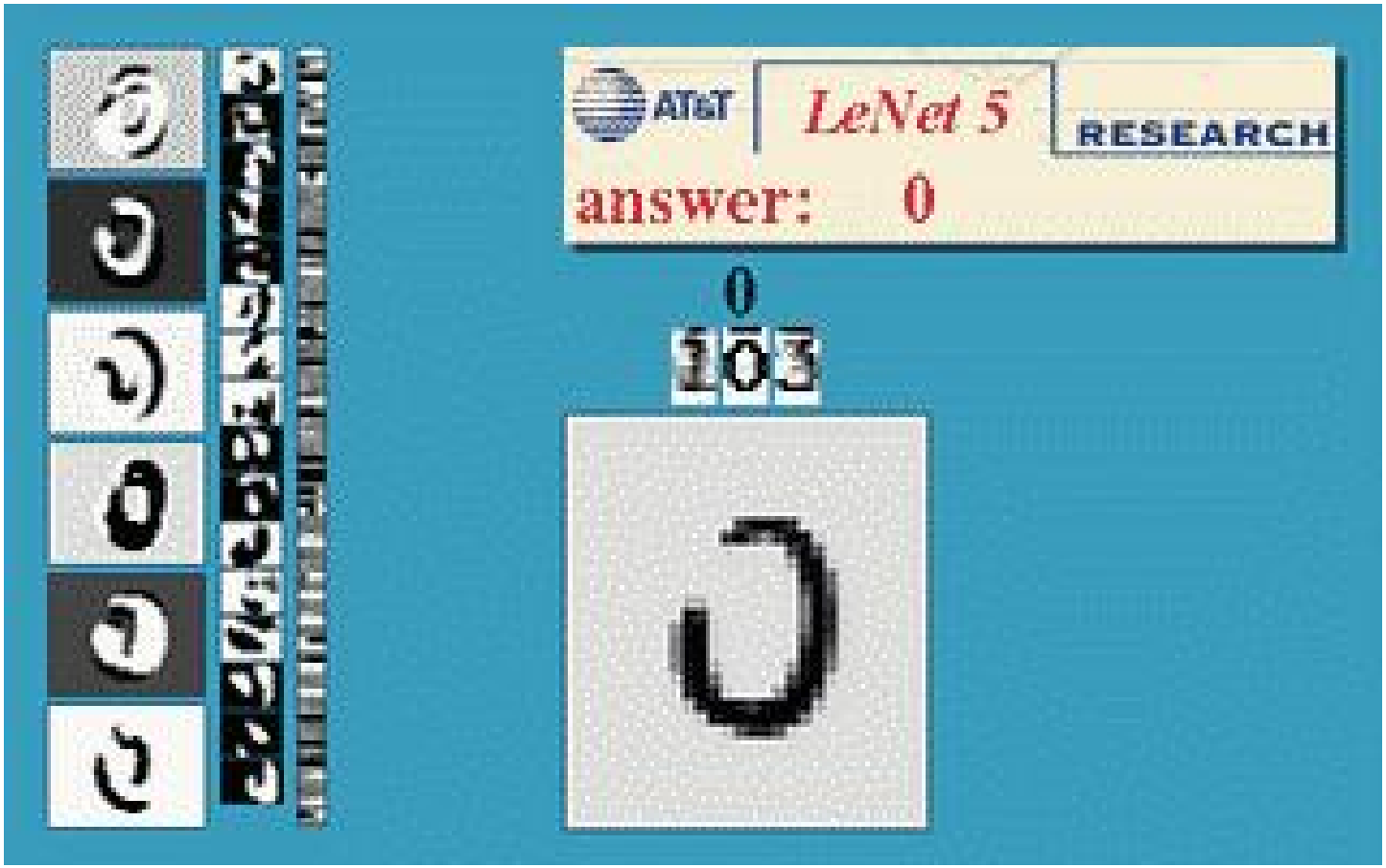
Drastically demonstrates that these systems are nothing like us.

SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

The Human Visual System (7)

Compare the previous features to a visualization of hidden layers in LeNet 5 (trained to recognize handwritten digits)!



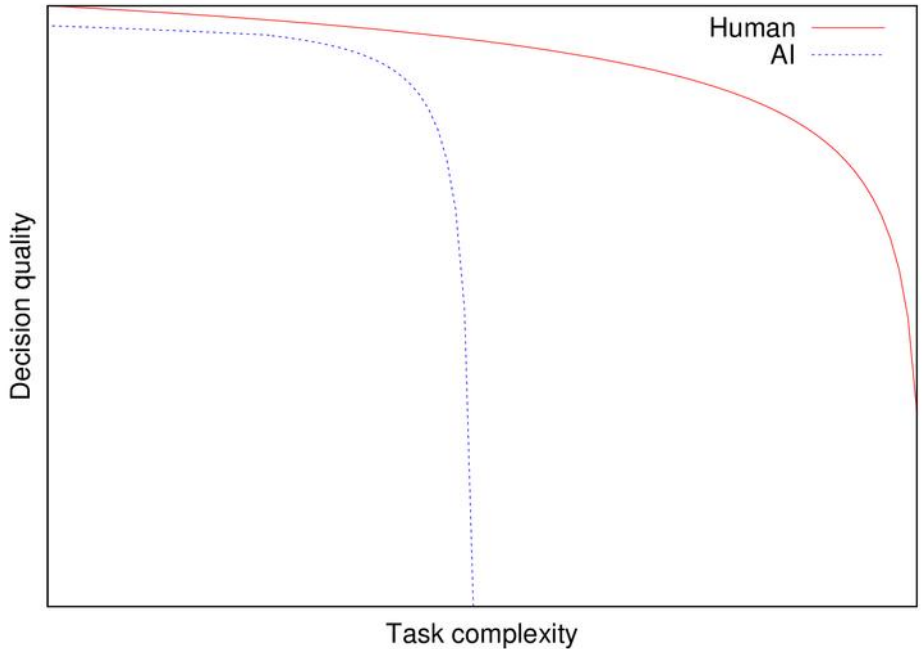
SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

AI Caveats

- AI errors are qualitatively different from human error (less generalization accuracy, "catastrophic failure" / no graceful degradation)
- Most successful systems (deep learning) are least understandable - very hard to characterize error modes!

- Very subtle correlations with outcome will be found by ML algorithms. This makes it hard to apply non-understandable algorithms to historical data



SEEWALD SOLUTIONS

alex@seewald.at
http://www.seewald.at

Ex.: Visual images from fMRI cortex activity

Reconstructing visual experiences from brain activity evoked by natural movies. Shinji Nishimoto, et al., Current Biology, published online September 22, 2011.

- Record brain activity using fMRI while viewing videos (several hours per person)
- Build statistical model to predict brain activity from video stream using machine learning techniques
- Apply to 18 million seconds of random videos (YouTube)
- Record brain activity for a different stream (90min)
- Recreate viewed video by averaging the 100 streams with the most similar activity - showing the best 30 seconds...

Video

- Mind's Eye (2023; using DL embeddings, just image reconstruction)
  https://medarc-ai.github.io/mindeye/

SEEWALD SOLUTIONS
alex@seewald.at
http://www.seewald.at