1 The Study of Intelligence

New Artificial Intelligence

- Sichtweisen: "Gehirn als Sitz der Intelligenz" ?
- Thematisieren, Methoden
- "es könnte auch anders sein"…

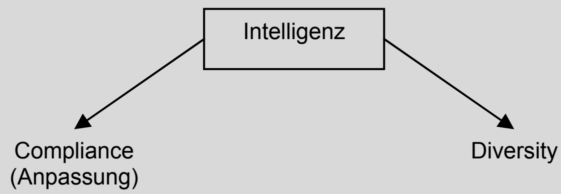
Intelligenz
Charakteristiken

(see p. 7-12)

- Auffassungsvermögen
- zielgerichtetes Handeln
- Lernfähigkeit
- Entscheidungsfähigkeit
- Analysefähigkeit
- Sprachvermögen
- Erinnerungsvermögen
- Wahrnehmung / Erkennen von Mustern
- abstraktes Denken
- Anpassungsvermögen
- Kreativität
- Beurteilungsvermögen
- Selbsterkennung

…

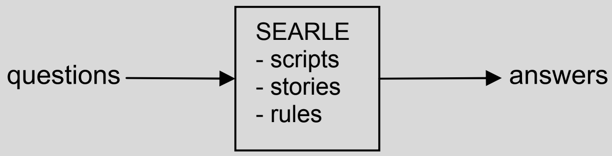
diversity-compliance trade-off



Compliance (Anpassungsvermögen) – Diversity (Verhaltensvielfalt)
complying with existing rules – coming up with something new

Allen Turing "Können Maschinen intelligent sein?"
Turing-Test: Versuchspers. am Computer kommuniziert mit Mensch/Computer. Muss herausfinden, ob Mensch oder Computer.
ELIZA: MIT

Searle:



terms

cognitive science ⇔ artificial intelligence (AI)
cognitive science: empirical sciences (psychology, biology, …)
AI: computer science, algorithms, logic

classical AI ⇔ NAI (embodied cognitive science)

NAI ⇔ GOFAI

"**N**ew" **A**rtificial **I**ntelligence ⇔ **G**ood **O**ld **F**ashioned **AI**

Embodiment (Verkörperung) – Mitte 80er Jahre, Weiterentwicklung

analytic ⇔ synthetic	***synthetic*** methodology: "understanding by building" – bottom-up Versuch, Verhalten zu Konstruieren, Nachbauen, Modellieren (e.g. Braitenberg-vehicles) ***analytic***: "understanding by separating into parts" – top-down geht von System aus. Versuch, Verhalten zu verstehen, durch Analyse Trend: synthetisch (dank Computersimulationen)

2 Foundations of Classical Artificial Intelligence and Cognitive Science

classical view	GOFAI **cognitivistic paradigm** (p. 39): Algorithms → cognition as computation → computation as operating on representations (Denken als Algorithmus/Programm) Turing-Machine
p. 36	Allen **Newell** / Herbert **Simon**: LT (logic theorist), thinking machine GPS (general problem solver) **George Miller**: Konzepte des Kurzzeitgedächtnis **7±2 "chunks"** (not bits!) e.g. Telefon# 00 49 89 3629064 **Noam Chomsky** – syntactic structures (natural speech)
PSSH	Physical Symbol System Hypothesis *A physical symbol system has the necessary and sufficient means for general intelligent action* (Newell and Simon, 1976) → symbol manipulation (empirical approach)
sense – think – act (p. 56)	• Denken • Problemlösen • Folgen
topics of classical AI (p. 46)	• knowledge representation • logic, planning • natural language processing • problem solving / reasoning • expert systems (rule based programs) • qualitative reasoning about physical processes • theorem proving • machine learning • computer vision • robotics
Turing machine (p. 41)	universal information processors abstract machine, universal, unlimited tape

the mind's main tasks (Johnson-Laird's "JL" Robot)	• perceive the world: vision • Learning, Memory, Action • think, create new ideas • control communication with others • feelings, intentions, self-awareness
design principles of Classical AI	c1) Model as computer program c2) Goal-based designs c3) Rational agents c4) Modularity c5) Sense-think-act cycle c6) Central information processing architecture c7) Top-down design
emergence (p. 124)	= behaviours **not preprogrammed** into the agents (that are surprising and not fully understood) behavior emerges from the **agent-environment interaction** Behavior can often be produced in **simple, cheap, and elegant** ways by exploiting **emergence**. Bsp: Roboter programmiert um Hindernissen auszuweichen. Als emergentes Resultat räumen sie jedoch auf. Roboter helfen einander, da sie ineinander stossen. → "Design for emergence" (Luc Steels)

3 *The Fundamental Problems of Classical Artificial Intelligence and Cognitive Science*

real world ⬄ virtual world

(p.61)

Das kognitivistische Paradigma vernachlässigt die Interaktion mit der realen Welt.

formale Welten (**virtuelle**)	**reale** Welten (nonformal)
Schach	*Fussball*
genau gegebene Züge (**Handlungen**)	Handlungen **nicht** vordefiniert
endliche Menge von Möglichkeiten nur genau **definierte Situationen**	viele **nicht** vordefinierte Situationen
vollständige Information	vollst. Information **nicht** definiert
Einflüsse von aussen **nicht** vorhanden	**Einflüsse von aussen**
"Schritte"	Zeit/Dynamik
sequentielle Ausführung	**parallel**

problems in classical AI (p.63)	• **Robustness and Generalization** • **Real-Time Processing** • **Sequential Nature of Programs** • Goal-based • hierarchically organized • process information centrally
Frame-Problem (≠F-O-R !!!) (p.65ff)	How can a system be **kept in tune with the continuously changing environment (real world)**? What is **relevant**, what is not? → Beziehung des Roboters zur realen Welt R1 robot, R1D1 robot-deducer, R2D1 robot-relevant-deducer Mensch weiss sofort, was relevant ist und was nicht. Er macht gewisse Annahmen, ohne sie explizit zu machen ("davon ausgehen"). Er muss sie machen, weil er sonst nicht überleben könnte. 1) Wagen bewegen → Bombe kommt mit (offensichtlich) 2) R1D1: versucht riesige Anz. von Seiteneffekten zu berücksichtigen. → fast alles irrelevant 3) R2D1: relevante von irrelevanten unterscheiden. Muss jedoch alles irrelevante "durchtesten" Schwerpunkte: → **Vorhersage-Problem (prediction)** → **Qualifikations-Problem (qualification)** → **Informationsrelevant**
symbol grounding problem (p.69f)	**How symbols relate to the real world.** Bsp. TRINKEN: - Flüssigkeit - Gefäss (keine Löcher, lässt keine Flüssigkeit durch, vernünftige Grösse) - Eigenschaften von Flüssigkeiten (Schwerkraft, Oberfläche horiz., schwappt über, giftig?, heiss/kalt, wasserhaltig, Geschmack,…) - Becher zum Mund führen (langsam) - Gewicht - schluckweise trinken - Erscheinungsbild / Bewegung / Lichtreflexion (gefroren/heiss) - Becher kippen - Durst: löschen - soziale Aspekte - Trinken: Blase → EMBODYMENT → sensory-motor grounding
"homunculus" (p. 73)	**homunculus problem** "little man" internal mechanisms

4 Embodied Cognitive Science: Basic Concepts

complete agent

- **adaptivity**
- **autonomy**
- **self-sufficiency**
- **embodiment**
- **situatedness** (Informationen über eigene Sensoren einholen)

self-sufficiency increases an agent's degree of *autonomy*
adaptivity is a consequence of *self-sufficiency*
situatedness is a prerequisite (Voraussetzung) for *autonomy*

Masanao Toda
"Fungus Eater"

→ autonomic, self-sufficient, situated, embodied

Bewegt sich in der realen Welt → Interaktionen, höchst komplex

Autonomie: - handeln ohne Anweisungen
- Unabhängigkeit
- Abhängigkeit von Umwelt
- Selbstständigkeit

F-O-R: Autonomie relativ zu anderen Akteuren / Umwelt, Eigenschaft einer Relation

adaption

(p. 93)

1) **evolutionary adaption**
   "peppered moth" (Schabe)
2) **physiological adaption**
   heiss → schwitzen
3) **sensory adaption**
   Pupillen → verändern
4) **adaption by learning**

ecological niche

Animals and humans are always "designed" by evolution for a particular ecological niche.

Cataglyphis (desert ant):
- path integration with polarized light compass
- visual landmark navigation (average landmark vector model) → matching, snapshot
- fixed search pattern

artificial agents

agent building goals:
- building for a particular task
- studying general principles of intelligence
- modeling certain aspects of natural systems

robotic agents ⇔ simulated agents

see p. 107

F-O-R frame-of-reference problem (p. 112)	F-O-R concerns the relation between observer, the designer, the artifact, the environment, and the observed agent. issues: ✓ **Perspective**: Beobachter (observer looking at an agent) Agent (agent itself) ✓ **Behavior-versus-mechanism** "Kategorienfehler" behavior is always the result of a system-environment interaction ✓ **Complexity** - scheinbar complex (Beobachter), jedoch einfacher Mechanismus - Komplexität in Umwelt
"Simon's ant on the beach"	Herbert Simon ("LT" – logical theorist) **"Simon's ant on the beach"** *Perspective:* To the ant, there are no pebbles, rocks, and puddles as we see them *behavior-vs-mechanism:* behavior must be clearly distinguished from internal mechanism complex path (observer) – internal simple rules: "if obstacle sensor on left is activated, turn right" *complexity:* the ant's behavior looks complex to an outside observer, but in fact it came about by applying simple rules. Komplexität kommt zustande aufgrund der Interaktion mit der Umwelt.
high level ⇔ low level specification	**high level specification** (domain ontology): specification of the intended behaviors systematic account – a list – of all the basic concepts that are needed in a particular domain. e.g. expert system, database system **low level specification**: specification of sensors, actors, component interaction, body morphology → the agent's physical setup → great room for interpretation the primitives that will be used in the design of the actual system
perspectives for explanations (p. 128)	levels / time perspectives: **short-term** perspective: explains why a particular behavior is displayed by an agent based on its current internal and sensory-motor state. → immediate cause of behavior **ontogenetic** perspective: resorts also to some events in the more distant past in order to explain current behavior. → learning and development perspective **phylogenetic** perspective: how the behavior evolved during the history of the species → evolutionary process (**ultimate or functional** perspective)

checklist agent experiments

a. research goal
b. tasks/desired behaviors
c. low-level specification
d. platform
e. control architecture
f. experimental setup
g. predictions / hypotheses
h. perform experiments
i. describe behavior
j. explanations of behavior

5 Neural Networks for Adaptive Behavior

Neuronen/Synapsen

Neuronen 10^{11}
Synapsen 10^{14}

NN characteristics

robust, can **learn** / **generalize**, can be **embedded**

Computer vs. Gehirn

Gehirn	***Computer***
viele rel. einfache Prozesse	**wenige rel. komplexe** Prozesse
weitgehend **parallel** (real-time performance)	**sequentiell**
Redundanz	wenig Redundanz
Robustheit → Fehler- und Rauschtoleranz → Generalisierungsfähigkeit	wenig Robustheit
analog	**digital**
embodied	unklarer Bezug zur realen Welt
Lernfähigkeit	--
vergessen	**vergessen nicht**

NETtalk

NETtalk, Sejnowski

text-to-speech: vorlesen
"Netzsprech"

natürliche Sprache äusserst komplex:
wave – gave – save – h*ave*

What means '**ghoti**' ?

enou<u>gh</u> gh
w<u>o</u>men o
na<u>ti</u>on ti

→ ghoti = <u>fish</u>

NETtalk:
Input – (gewünschter) Output bekannt
→ **supervisiertes Lernen**

4 or 5 basics

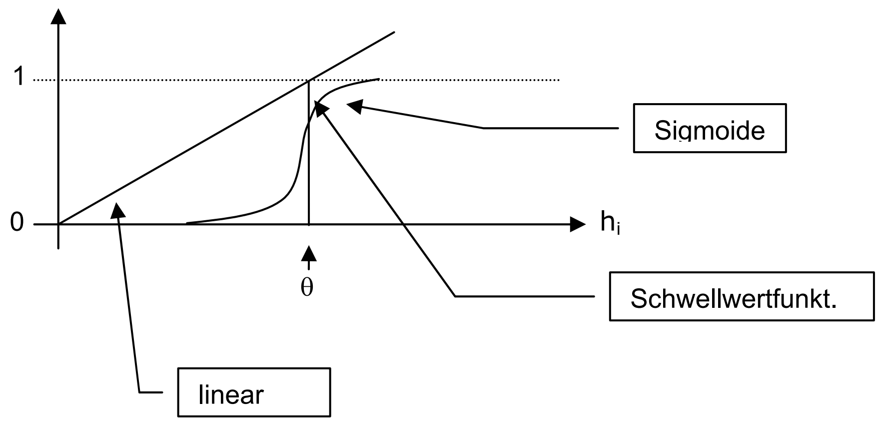
(p. 146ff)

- **characteristics of the node**

$$h_i = \sum w_{ij} o_j$$

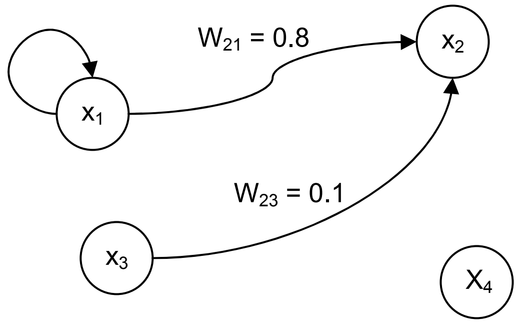
$$Aktivierungsfunktion \quad a_i = g(h_i)$$

$$\theta_i : Schwellwert$$



- **connectivity**

wie sind die Layer miteinander verbunden?



	x_1	x_2	x_3	x_4
x_1	1			
x_2	0.8	0	0.1	
x_3			0	
x_4				0

- **propagation rule**

diskrete Zeitschritte $h_i = \sum w_{ij} a_i$

- **learning rule**

Hebb'sche Lernregel

$$\Delta w_{ij} = \eta \cdot o_j \cdot a_i$$

$$\eta : Lernrate$$

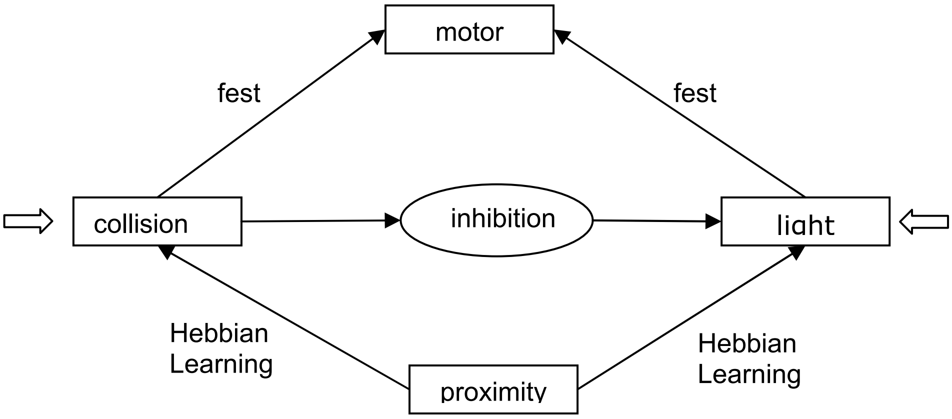
$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}$$

- **embedding the network in the agent**

Neural Net is embodied

DAC

C-DAC
Complete Distributed Adaptive Control architecture



learning types

supervisiert vs. nicht-supervisiert

(p. 167)

Nicht-supervisiertes Verfahren (Selbsorganisation)
- Hebb'sches Lernen (→ DAC)
- Kohonen-Netzwerke (Clustering) p.167

Supervisiertes Verfahren
Input → gewünschter Output
NETtalk
(MLP: multi-layer-perceptrons, Backpropagation)

Reinforcement Learning
true/false
credit assignment / blame assignment problem

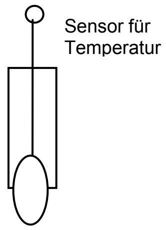
6 Braitenberg Vehicles

14 Types
built in a bottom-up fashion

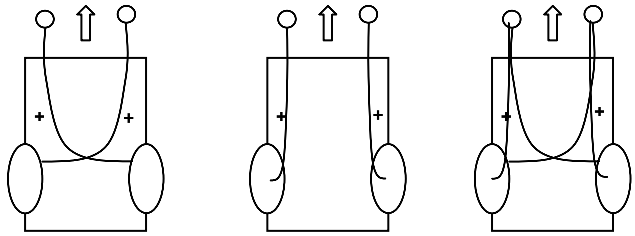
Type		Sensors	Motors	Quality
1	*getting around*	1	1	only activation
2	*approach and avoidance*	2	2	only activation
3	*attraction*	2n	2	+ **inhibition** (+ multi-sensorial)
4	*values and special tastes*	2n	2	+ **nonlinear sensor-motor** coupling + tresholding
5	*brain power*	2n	2	+ **brain power** + learning
6	*evolution*	2n	2	+ **evolution** + neural networks
7-14		-	-	(cognitivistic flavour)

types 1-4

vehicle 1:

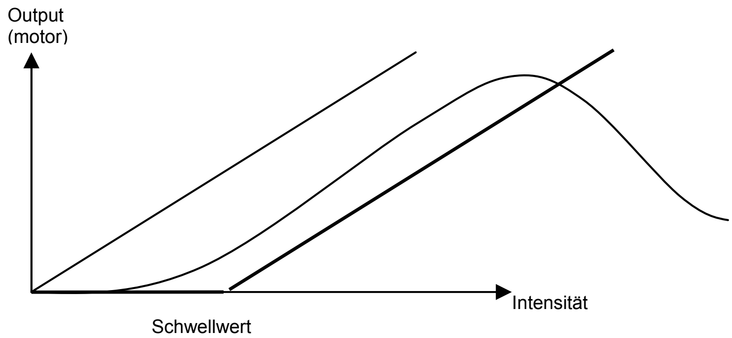


vehicle 2:



vehicle 3: + / - (Inhibition)

vehicle 4:



perspectives

- **Frame-of-reference problem**
- **behavior segmentation & action selection**
- design of autonomous agents
  (EBA – extended Braitenberg Architecture)

type 6: evolution

"survival of the fittest" – Die nicht überlebensfähigen verschwinden. Die besten warden ausgewählt.
→ cumulative selection
Richard Dawkins: "The blind watchmaker" (Evolution)

segmentation of behavior

(p. 195)

e.g. vehicle type 3c, moves toward light-source
What action is it involved in, "turning toward the light" or "turning away from the obstacle"? (view of observer)
→ dynamics of internal variables in interaction with environment
EBA: architecture extended by internal variables

law of uphill analysis and downhill invention

building something new (invention) is easier than **analyzing something that already exists** (analysis)!

ecological balance

"principle of ecological balance"
z.B. ware menschliches Hirn in vehicle 1 sinnlos

7 The Subsumption Architecture

tasks

- aspects of **natural evolutionary theory**
- simple, sensor-action couplings with **little internal processing**
- **intelligence not centralized** (brainlike), emerges from loosely coupled processes

→ **Parallelität**: Prinzip der parallelen lose gekoppelten Prozesse (nur Inhibitions-links)
→ **Sensor-Motorische Koordination**:
Feedback von der Umgebung
Erzeugen von Sensorstimmulierung, die beinflusst, was man tut
John Dewey (1896) "sense-think-act cycle" (think: innere Repräs.)

"behavior-based approach"

→ subsumption-based
→ whole field of embodied cognitive science
→ "non-information-processing-based" ⇔ "knowledge-based"

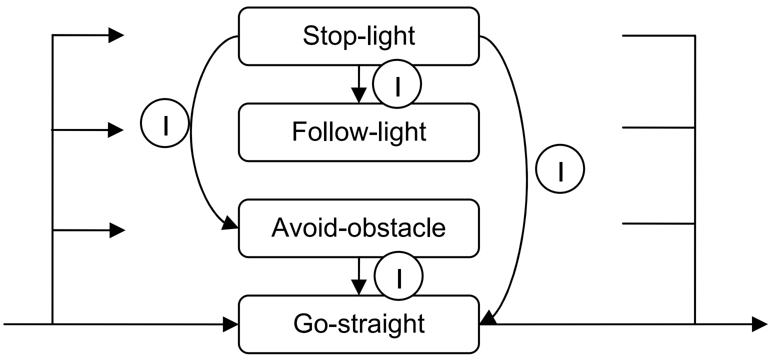
competences (functional decomposition)

Rodney Brooks (MIT):

S – M – P – A cycle
[sense – model – plan – act]

klassische Sichtweise: **S – T – A** cycle (sense – think – act)

subsumption



Eine Schicht hat Kontrolle → kompetitives Schema (action selection!)

Immer alle Module aktive, aber nur eines ist der "Chef"

FSM

(p. 205)

FSM: Finite State Machine
A-FSM: augmented FSM

Input \ Zustand	vorwärts **1**	stop **0**
IR > θ	0	0
IR < θ	1	1

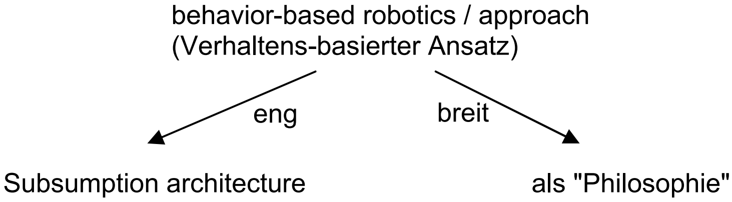
Zeit nicht massgebend in FSM
falls Zeit massgebend, z.B. Roboter Rechteck → A-FSM

3-constituents principle

- **ökolog. Nische**
- **gewünschtes Verhalten (desired behavior)/Aufgaben (tasks)**
- **Akteur**

Braitenberg (Sensorstimulat. zentr.) ⇔ Subsumption

Terminologie



8 *Artificial Evolution and Artificial Life*

evolution examples

- **Braitenberg type 6 vehicles**
  vehicles running around on a table, copying vehicles with possible errors, some vehicles drop on the floor → Darwinian evolution
- **autonomy**
- **subsumption architecture**

künstliche Evolution: "life-as-it-could-be"

"the blind watchmaker"

- ✓ Evolution hat kein Ziel
- ✓ Lösungen immer im Augenblick beurteilt
- ✓ Grundbausteine

Artificial Life definition

The study of man-made systems that exhibit behaviors characteristic of natural living systems. Synthesizing life-like behaviors within computer and other artificial media.

random selection vs. cumulative selection

random selection:
example of the monkey typing a Hamlet phrase:
("METHINKS IS LIKE A WEASEL")
http://home.pacbell.net/s-max/scott/weasel.html
→ ungeeignet, auch für einfache Probleme

cumulative selection:
- Population
- Selection (Qualitätskriterien)
- Mutation (Lösungen verändert)

random selection (27^{28}) ⇔ **cumulative selection** (~43 generations)

cumulative selection: good properties accumulate rather than get lost in a random process.

genotype def.	The set of all genes is called the organism's ***genotype***. **genome** = a description of an agent's features (color of its hairs, eyes, skin; body size; limb size; shape of nose, …) Genotyp: Was auf Nachfahren übertragen wird Genom: "Datenträger"
genotype vs. phenotype	**genotype** (genetic setup) ⇔ **phenotype** (final organism) Through the process of development, the genotype is translated into a phenotype. Phenotyp: Verwirklichung des Genotyps
processes of evolution	(encoding) → **development** → **selection** → **reproduction** **selection:** taking the best individuals ignoring the others → *loss of diversity !!* getting just the right mix of individuals → *exploration-exploitation trade-off* (search good individuals but still explore other regions) solution: → *roulette wheel selection* **reproduction:** crossover & mutation

9 Other Approaches

10 Design Principles for Autonomous Agents

design principles

(p. 303)

meta-principle

Constituents of design (meta-principle)	
1	The three-constituents
Morphology, architecture, mechanism (→ agent itself)	
2	complete agent
3	parallel, loosely coupled processes
4	sensory-motor coordination
5	cheap design
6	redundancy
7	ecological balance
8	value

Grosse Demütigung:

- Kopernikanische Wende (Erde nicht Mittelpunkt)
- Darwin'sche Evolutionstheorie (Mensch stammt ab von Affen)
- Genetischer Code / Doppel-Helixe (Reproduktion des Menschen gleich wie Hefepilz)
- Gene von Affen 98%
- **Intelligenz**: einfache Maschine