Framsticks mind experiments

Maciej Komosinski

www.framsticks.com

This presentation
is based entirely
on the works of
prof. Pete Mandik [Man03]
Cognitive Science Laboratory
Department of Philosophy
William Paterson University of New Jersey

Philosophy: core question

- what is the relation of the **mind** to the world
- . . . such that the mind has representations of the world?

materialistic view:

- how brains (physical systems) have representations of the world?

What is representation?

- a state of an organism (its brain) that carries information about environmental and bodily states (Dretske 1988; Milikan 1984; 1993)
- discussion: information, isomorphism, encoding, decoding

The question in two versions

- synchronic
  - what patterns of structure and activity in the world support the representation of objects, properties, and states?
- diachronic
  - what happened over time for physical structures to have representational contents?

Neurosemantics

- what it means for states of NNs to have representational contents?
- related to:
  - philosophy of neuroscience
  - philosophy of mind

“Having neurons, build mind”

- what would you ask for?
  - more neurons? (complex brain?)
  - body? (embodiment?)
  - evolutionary mechanisms?
  - knowledge about required mind states/representations?
- what is the *simplest set of conditions* for the implementation of mental representations in NNs?

Diachronic approach: temporal and casual priority of representations and nervous systems

- NSs existed before or after organisms with representational states?
- did NSs evolve **in order to** provide the means for representing?
- or did they serve some non-representational function first?

AI and computational neuroscience

- synthetic approach
- **synchronic** aspects of the problem of representation
- construction of NN controllers
- testing hypotheses about possible NN architectures supporting intelligent behavior
- but...

Diachronic aspects not tested

- how might the proposed NN architectures have evolved from other systems?
- artificial life techniques are proper for such questions!

Differences

GOFAI (good, old-fashioned AI)

- modeling simplified subsystems of agents
- focused on designed aspects

Artificial Life

- modeling simplified agents
- focused on evolved aspects
- holism

Artificial life so far

- focused on evolving "minimally cognitive behavior"
  - obstacle avoidance
  - food finding
- often in opposition to the assumption that *intelligent behavior equires mental representation and computation*

Representations not needed?

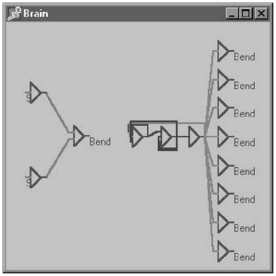
- Rodney Brooks: "representation is the wrong unit of abstraction in building . . . intelligent systems"
- Randal Beer: "the design of the animat's nervous system is simply such that it . . . synthesizes behavior that is appropriate to the . . . circumstances"

Reactive agents!

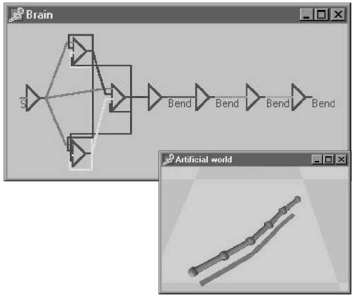
- surprising variety of behaviors
- in spite of lack of internal representations of environments
- very simple agents exhibit minimally cognitive behavior
- agents do not need internal states, so there is no inputs/outputs transformation, no computation, and no representation

Control systems in food finders (positive chemotaxis)

- modular – two control systems:
  - locomotion (for example CPG)
  - spatial location of the stimulus



- non-modular
  - swims around continuously in wide curved arcs
  - smell sensor active: arcs become tighter, food is absorbed
  - smell sensor high activity: CPG stops



Representation vs. Modularity

modular

- position of food
- 2D: near–far, left–right
- decoded by the single turning muscle

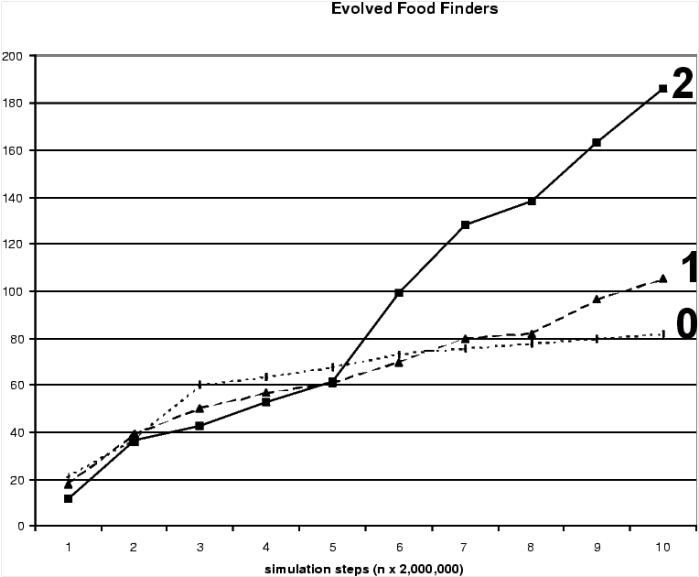
non-modular

- proximity of food
- 1D: near–far
- decoded by muscular system $\rightarrow$ curvature of arcs

Advantages of representation

- optimization: identical bodies, similar NNs
- fitness: ability to find food
- optimized: NN weights only
- smell sensors: none, one, or two
- NN topology: outputs for muscles (9), inputs for smell sensors, 2 hidden layers with all feed-forward and (for one layer) all feedback connections
- averaged from five evolutionary runs per creature

Experimental results



Number of smell sensors

Conclusions

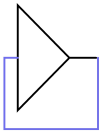
- simple ANNs are capable of supporting representations of spatial locations of stimuli
- Alife as the way of creation of thought experiments of indefinite complexity (Dennett, 1998)
- can we build a gradualist bridge from simple amoeba-like automata to highly purposive intentional systems, with goals, beliefs, etc.? (Dennett, 1998)
- representational and computational systems will figure very early in the evolutionary trajectory from mindless automata to mindless machines (Mandik, 2001)

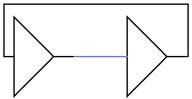
Four categories of mobile animats

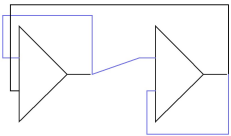
- Creatures of Pure Will (no sensory inputs)
- Creatures of Pure Vision (perceive environment)
- The Historians (memory mechanism)
- The Scanners (comparison of environment and internal states)

Creatures of Pure Will

- synthetic psychology and synthetic neuroethology: what are the simplest systems that exhibit mental phenomena?
- common assumption: movement required
- repetitive signals
- CPGs

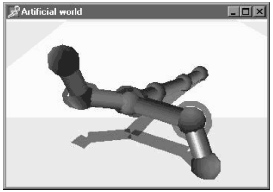


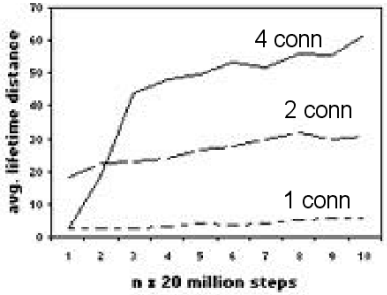




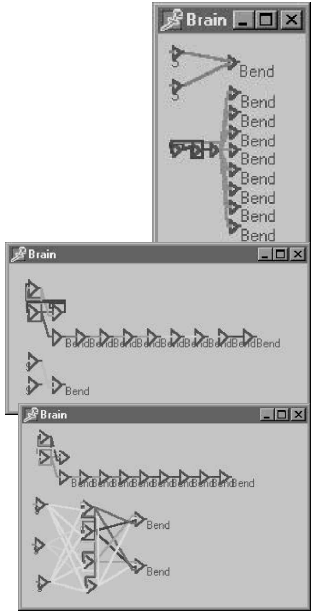
Comparison

- are more complex CPGs advantageous?
- constant body
- three kinds of CPGs
- motor imperative (procedural) representations *can be* the product of evolution without indicative (declarative) sensory input!



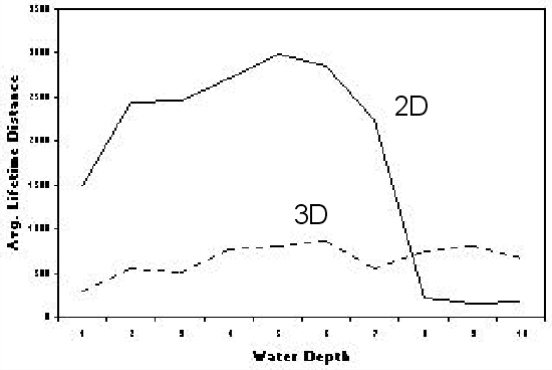


- taxis: toward/away from stimulus (ex. positive phototaxis)
- kinesis: motion triggered/suppressed by stimulus (ex. running within some temperature range)
- CPGs + orientation neurons



2D/3D sensing in water

- it is not *always* good to represent more.

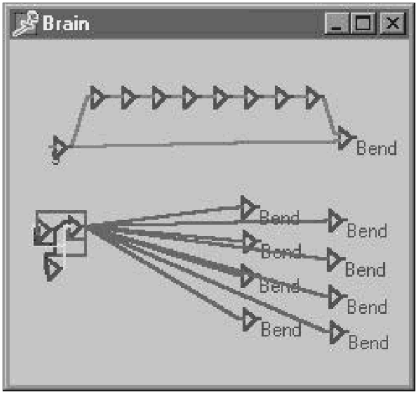


The Historians

- short-term memory: recurrent NNs
- memory = encoding, maintenance, retrieval
- retrieval: how to utilize stored information?
- but the problem is known in nature: *E. Coli*
  - so small that it cannot use multiple sensors
  - but has to determine the direction of greatest concentration of nutrient
  - it memorizes concentration (internal states!)
  - changes heading when lower concentration detected

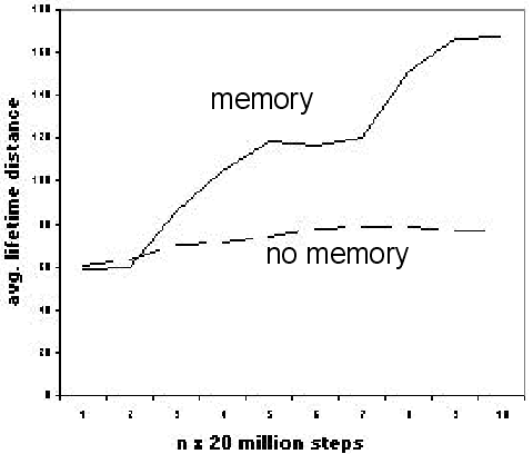
Memory implementation

- 8-neuron propagation delay
- representation of spatial distance from stimulus
  - no memory: two sensors (difference)
  - memory: single sensor!



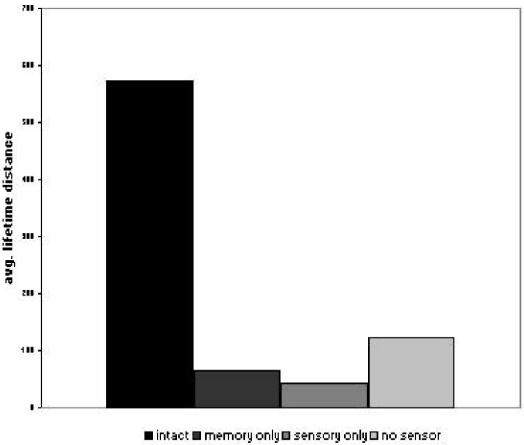
Memory vs. no memory

- The Memorians construct representation of 2D location of food: **unknown**
- The Memorians utilize representation of the past: **sure** (encode, maintain, retrieve)



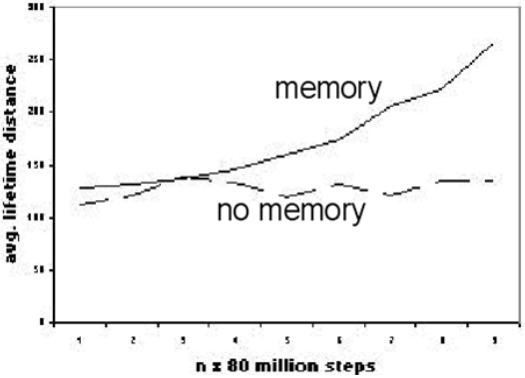
Question #1

- do they compare delayed (memory) and current (perception) signal, or is the delayed signal only useful?
  - verification: memory buffer works (all weights nonzero)
  - removal of some connections



Question #2

- if delay buffer weights set to 0, will they evolve to be non-zero?
  - **yes.**

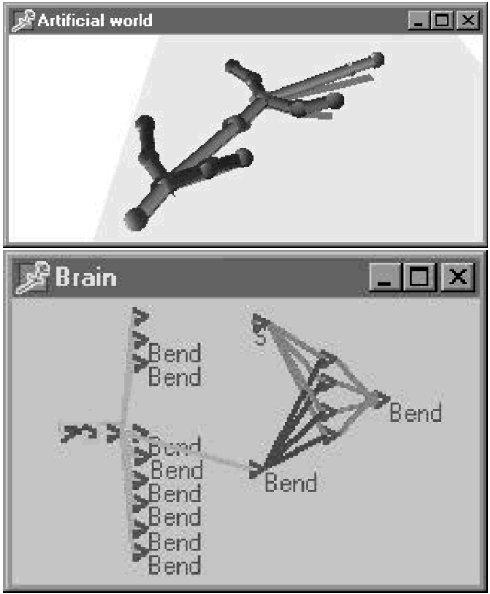


Behavior of The Historians

- pirouette motion similar to *C. Elegants* worms, which use gradient navigation
- . . . nematodes use similar kind of memory to the one evolved in Framsticks?

The Scanners

- radar: single sensor mounted on a long limb, used as an oscillating scanner
- CPG to control movement and radar position
- orientation muscle controlled by a NN with a sensor and radar position information



2D stimulus location

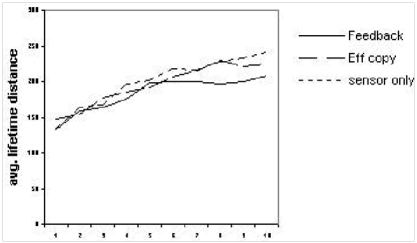
- smell sensor: active – food is where radar is directed
- smell sensor: inactive – food is elsewhere

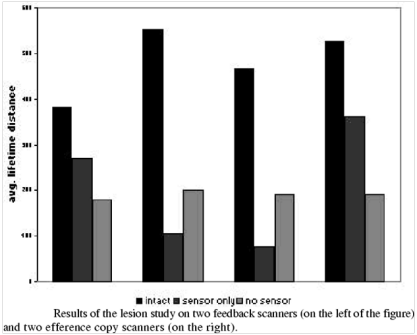
or

- correlation of smell sensor activity and radar control command

Comparison

- similar performances
- similar behavior?
  - all used only smell sensor?
  - all used all the information available? **yes:**
    - non-zero weights
    - lesion study





Results of the lesion study on two feedback scanners (on the left of the figure) and two efference copy scanners (on the right).

Evolved representations

	are	*of*
Pure Will	motor commands from CPGs	patterns of muscular movement
Pure Vision	states of sets of sensory transducer neurons and signals they passed to orientation muscles	current egocentric locations of food sources in 1D/2D/3D
Historians	as Pure Vision, plus memory buffers	as Pure Vision, plus past locations of food
Scanners	as Pure Vision plus Pure Will	as Pure Vision plus Pure Will

References to philosophical theories and paradigms

- teleological informational approach (Dretske 1995)
- isomorphism approaches (Cummings 1996)
- temporal evolution of neurosemantics (Millikan 1996)
- egocentric/allocentric representations

Critics

- simulations too constrained?
  - no, although more sophisticated scenarios will be useful. Still much to be done in simple simulations.

- simulations are mere simulations!
  - abstract from real phenomena and may leave out crucial features
    - a danger not peculiar to computer models, but all models.
  - computer simulation is not real, so it is virtual, and thereby *fictional*
    - nothing is real in computers!? no, computers are material.

References I

[Man03] Pete Mandik. "Varieties of Representation in Evolved and Embodied Neural Networks". In: *Biology and Philosophy* 18.1 (2003), pp. 95–130. URL: http://www.framsticks.com/files/common/Mandik_RepresentationsInNeuralNetworks.pdf.