What Kind of Models should be used for

Representing Cognitive Phenomena in Biological Systems

Joanna J. Bryson

Harvard Primate Cognitive Neuroscience Laboratory

jbryson@wjh.harvard.edu

http://www.ai.mit.edu/~joanna

Outline

- Modularity and Attention (in Philosophy and AI)
- Their Relationship to Biological Minds
- The “C” Word

Motivating Modularity

- Combinatorics is the Problem.
- Search is the only Solution.
  - Planning and Learning
  - Design and Evolution

The Task of intelligence is to focus search.

(bias, constrain)

Combinatorics and Search

- If . . .
  - an agent knows 100 actions (e.g. eat, drink, sleep, step, turn, lift, grasp, poke, flip, read . . .), and
  - it has a goal (e.g. go to Paris),
- Then . . .
  - Finding a one-step plan may take 100 tries.
  - Finding a two-step plan, may take 100^2 ($10,000$).
  - For a plan of unknown length, agent may search forever, missing a critical step or sequence.

Modularity

- Modularity simplifies design.
  - Decomposes the problem into simpler units.
  - Focuses search by using locally optimal representations (planning, learning, perception).
- It also generates design issues.
  - Decomposition
  - Coordination
  - Learning

Modularity: Philosophy to AI

- Fodor (1984): Modularity in perception, centralized thinking and action.
- Behavior-Based AI (Brooks 1986, Maes 1991): Modules encapsulate action *and* perception, distributed arbitration.
- Hybrid / Layered Achitectures (Georgeff & Lansky 1987, Gat 1992, Banasso et. al 1998): centralized decision making and learning, distributed action with some perception.

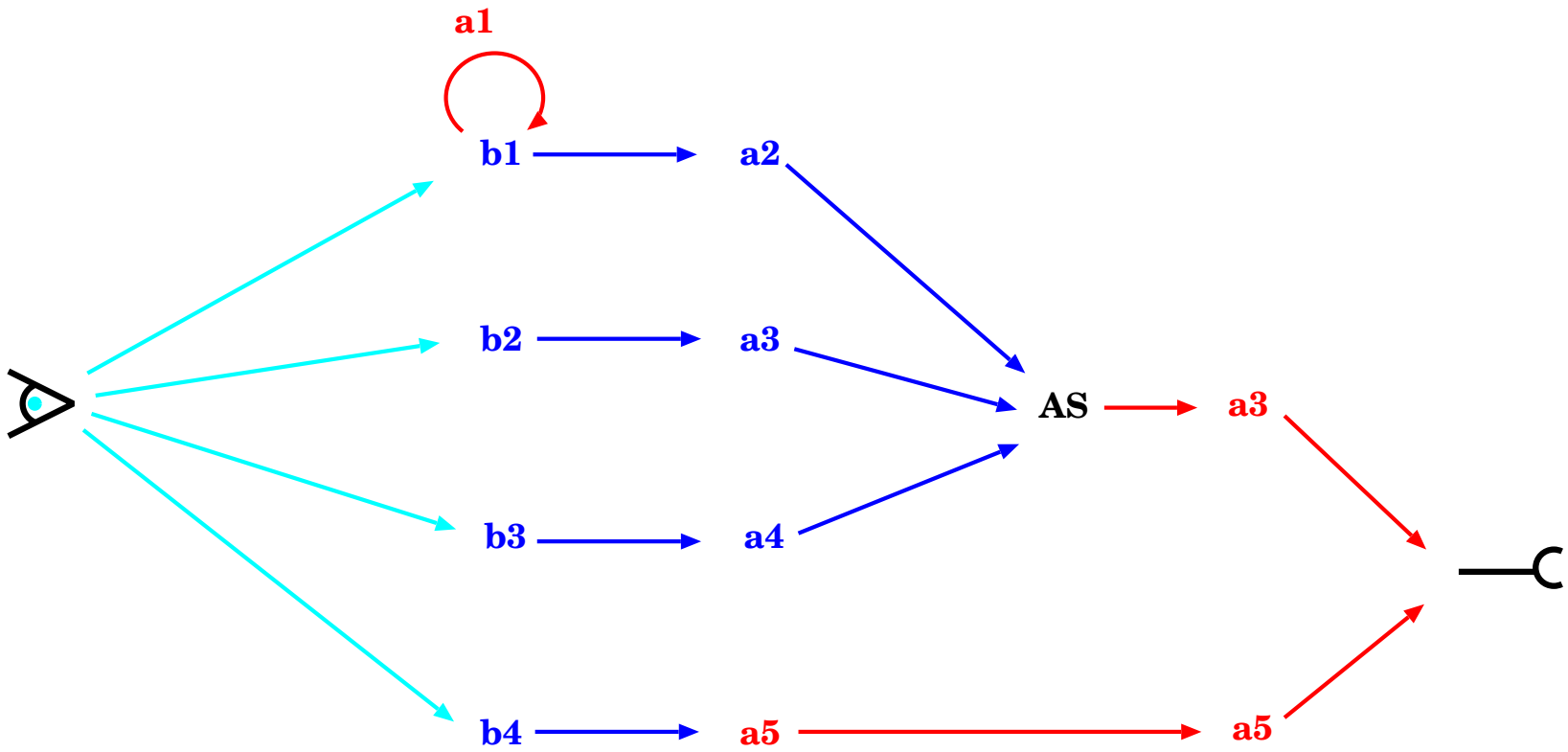
Behavior Oriented Design

(Bryson 2001)

BOD encapsulates perception, action *and* memory. Arbitration is itself a module. problems:

- Learning is done within modules.
- Modular decomposition is dictated by variable state.
- Coordination between modules is done by hierarchical reactive plans.

Behaviors and Behavior Arbitration in BOD



What is a Behavior? (in BOD)

- A module in a complete, complex agent.
- Control for agent's actions (expressed and/or internal).
- Perception required for that control.
- Memory required for perception or control.

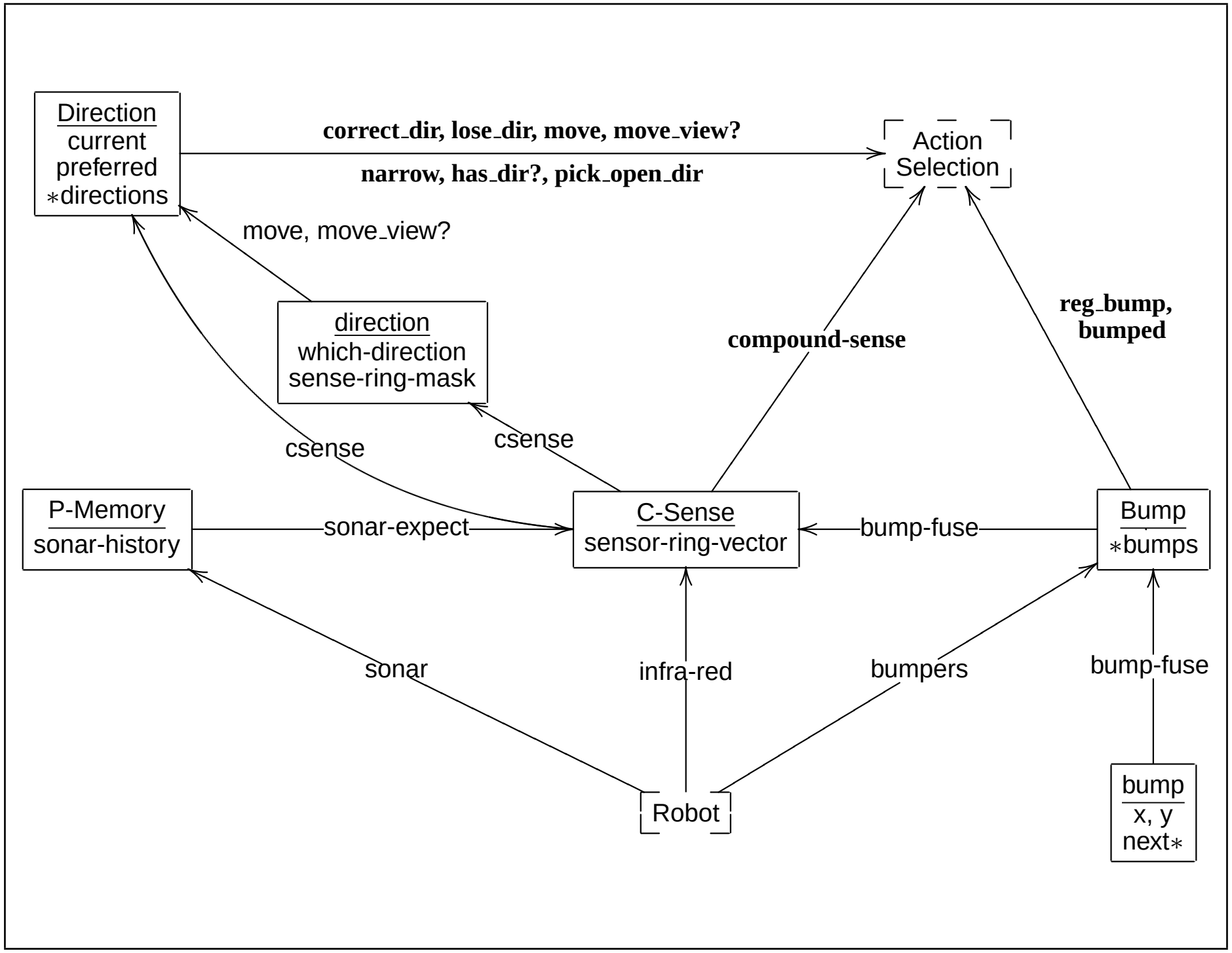
BOD Extends Behavior-Based AI to be more like Object-Oriented Design. (Parnas 1985, Coad 1997)

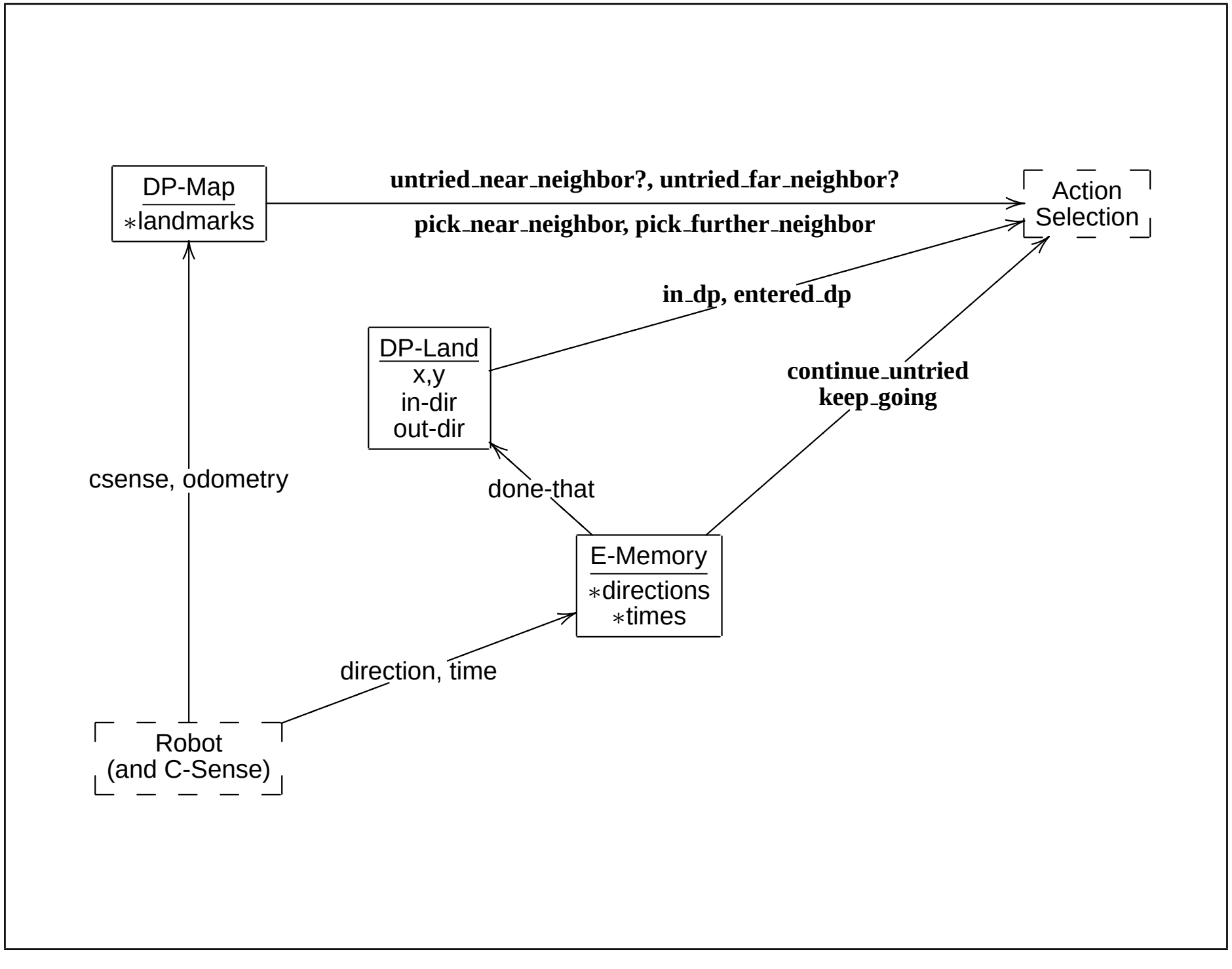
Reactive Plans in BOD

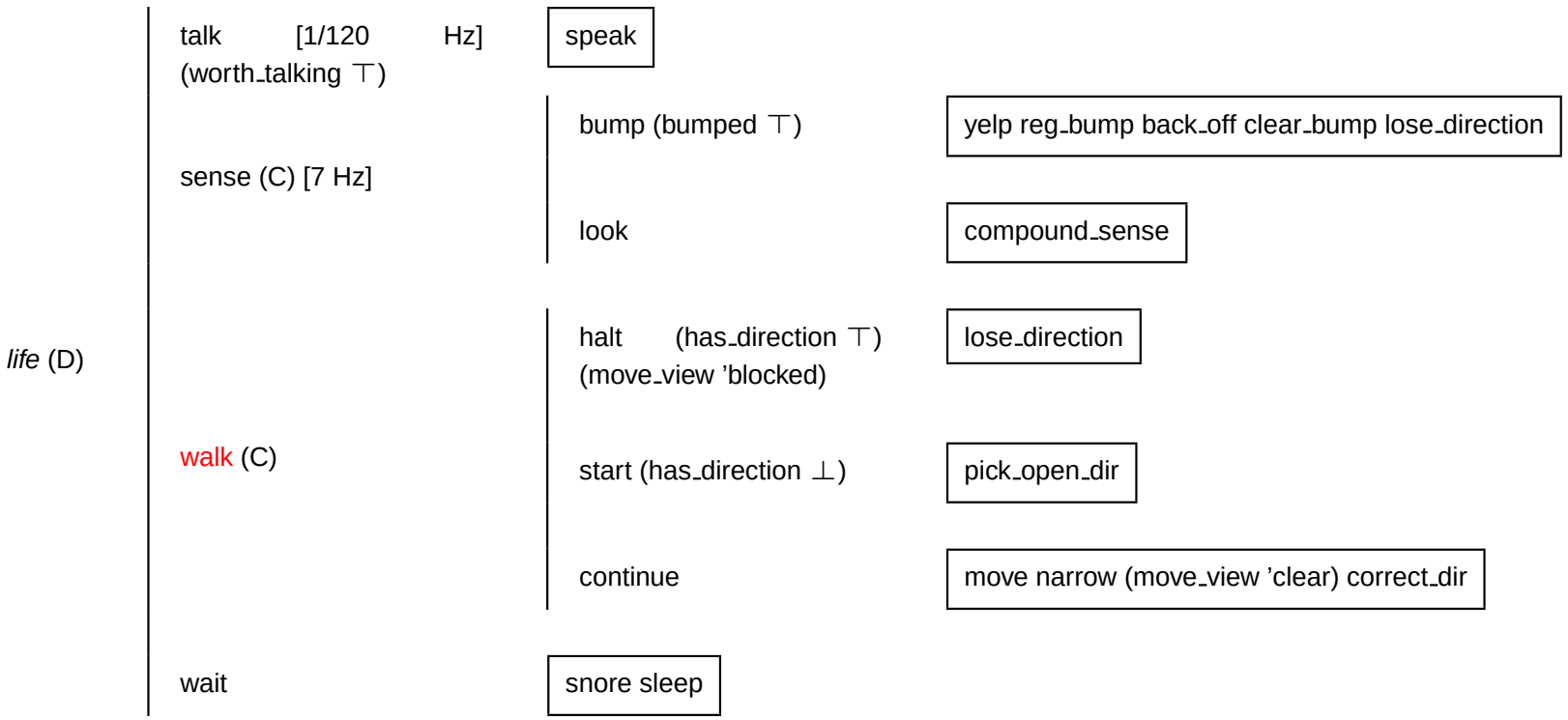
- Reactive plans are an engineered solution.
  - Planning
  - Reactive Planning
  - Reactive Plans
- Arbitration must address three situations.
  - Some things need to be checked at all times.
  - Some only consideration in particular contexts.
  - Some things reliably follow from others.

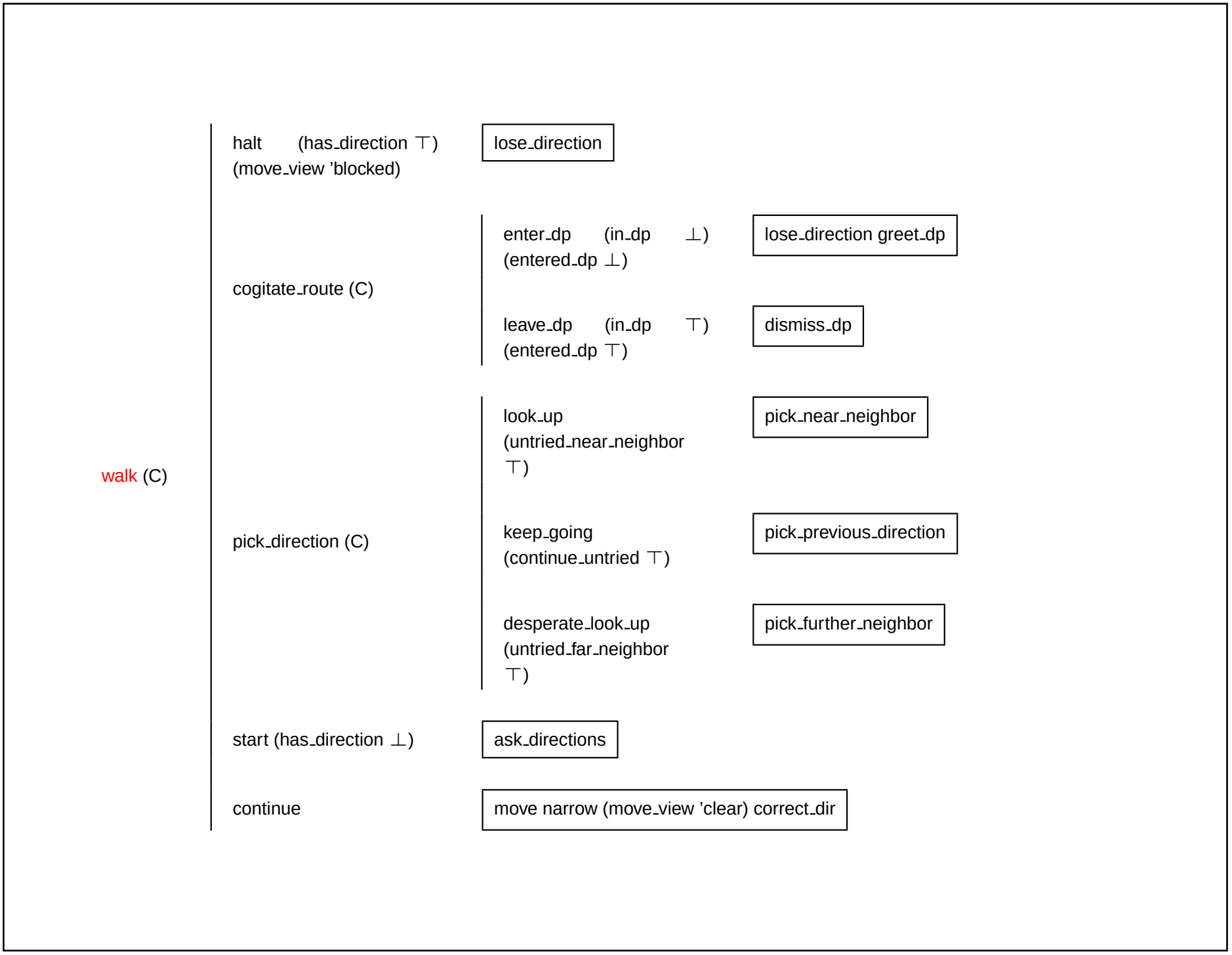
Example: A Mobile Robot

(Bryson ATAL97)


Direction
current
preferred
∗directions
correct_dir, lose_dir, move, move_view?
narrow, has_dir?, pick_open_dir
Action
Selection
move, move_view?
direction
which-direction
sense-ring-mask
compound-sense
reg_bump,
bumped
csense
csense
P-Memory
sonar-history
sonar-expect
C-Sense
sensor-ring-vector
bump-fuse
Bump
∗bumps
sonar
infra-red
bumpers
bump-fuse
Robot
bump
x, y
next∗


DP-Map
∗landmarks
untried_near_neighbor?, untried_far_neighbor?
pick_near_neighbor, pick_further_neighbor
Action Selection
in_dp, entered_dp
DP-Land
x,y
in-dir
out-dir
continue_untried
keep_going
csense, odometry
done-that
E-Memory
∗directions
∗times
direction, time
Robot
(and C-Sense)


life (D)
talk [1/120 Hz]
(worth_talking ⊤)
speak
sense (C) [7 Hz]
bump (bumped ⊤)
yelp reg_bump back_off clear_bump lose_direction
look
compound_sense
walk (C)
halt (has_direction ⊤)
(move_view 'blocked)
lose_direction
start (has_direction ⊥)
pick_open_dir
continue
move narrow (move_view 'clear) correct_dir
wait
snore sleep


walk (C)
halt (has_direction ⊤) (move_view 'blocked)
lose_direction
cogitate_route (C)
enter_dp (in_dp ⊥) (entered_dp ⊥)
lose_direction greet_dp
leave_dp (in_dp ⊤) (entered_dp ⊤)
dismiss_dp
pick_direction (C)
look_up (untried_near_neighbor ⊤)
pick_near_neighbor
keep_going (continue_untried ⊤)
pick_previous_direction
desperate_look_up (untried_far_neighbor ⊤)
pick_further_neighbor
start (has_direction ⊥)
ask_directions
continue
move narrow (move_view 'clear) correct_dir

Outline

- Modularity and Attention (in Philosophy and AI)
- Their Relationship to Biological Minds
- The “C” Word

Modules in Animals

- What are modules about?
  - Specialized representations.
  - Specialized computations / connectivity.
  - Differential time courses.
- How are they represented in brains?
  - Regions in the cortex.
  - Specialized organs.
  - Temporally differentiated configurations.

BOD's Biological Plausibility

(Bryson EmerNet00)

- Elements in BOD
  - Specialized Behaviors: Neocortex
  - Action Sequencing: Basal Ganglia, Periaqueductal Gray Matter
  - Attention Shifting: Amygdalic System
- Other Possible Additions
  - Smoothing and Interpolation: Cerebellum
  - Episodic Memory, Indexing, Rule Manipulations: Hippocampus, Enthorinal Cortex

Natural Intelligence Simulations

- Hierarchical Action Selection (Bryson 2000)
- Transitive Inference in Squirrel Monkeys (Bryson 2001)
- Competing Learning Systems in Cotton-Top Tamarins (in progress w/ Marc Hauser)
- Social Organization in Various Species of Macaques (in progress w/ Jessica Flack)

Outline

- Modularity and Attention (in Philosophy and AI)
- Their Relationship to Biological Minds
- Consciousness

Recommended Literature

- Memory sequencing / “Multiple Drafts” (Dennett & Kinsbourne 1992).
- Interrupt-driven help for automatic action sequencing (Norman & Shallice 1986).
- Lots of top-down attention / resources (Dehaene et al. 2001).
- Short Term vs. Long Term Memory (McClelland, McNaughton, O’Reilly 1995)

My Current View

- “Conscious spotlight” is extra activation.
  - Helps iterate through top-down / expectation options looking for sufficiently active path / mental context.
  - Can negatively interfere with some automatic processes.
- Recent episodic memory formed (& ordered) via indexes w/ “conscious spotlight”...

My Current View

- “Conscious spotlight” is extra activation...
- Recent episodic memory formed (and ordered) via indexes with “conscious spotlight”.
  - Recent EM integral to semantic and long-term episodic memory.
  - Indexes critical to some reasoning and learning, e.g. transitive inference, task learning.
  - Words as indexes (Spelke *in preparation*).
  - May not be continuously “conscious”, *or* spotlight may have boring default location.

Summary

- Modularity and Attention (BOD was my PhD)
- Their Relationship to Biological Minds
- The “C” Word

Joanna Bryson

jbryson@wjh.harvard.edu
http://www.ai.mit.edu/~joanna

[Talk Boundary]

Who am I and What do I Do?

Engineering: I build complete, complex agents which behave, learn and plan.

Science: I build and test models of animal cognition.

- Complete agents: artifacts with identity

  e.g. robots, VR characters, tutors, monitors, artificial life.

- Complex agents:
  - multiple, potentially conflicting goals, and
  - multiple, mutually-exclusive means of achieving those goals.

Artificial and Natural Intelligence

Artificial Intelligence has two goals:

- **Engineering** — Trying to improve computer science by emulating natural intelligence.
- **Science** — Trying to understand natural intelligence by designing and testing models.

Any piece of AI research should address *one* of these.

Artificial and Natural Intelligence

Both fields share some challenges, therefore they may also share some solutions.

- Limits in storage capacity, processing time.
- Characteristics of information environment

  e.g. light, sound, gravity.
- Combinatorial complexity of planning, learning.