Carving the world at its joints

Rich Sutton
RL&AI Laboratory
University of Alberta

Themes

- Carving the world at its joints
- Finding structure in the blooming, buzzing confusion
- Orienting AI around experience
- The computational theory of knowledge

Marr's three levels

at which any information processing system can be understood

- Computational Theory Level
  - What are the goals of the computation?
  - What is being computed?
  - Why are these the right things to compute?
  - What overall strategy is followed?

What and Why?

- Representation and Algorithm Level
  - How are these things computed?
  - What representation and algorithms are used?

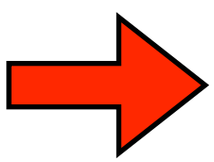
How?

- Hardware Implementation Level
  - How is this implemented physically?

Really how?

Marr's three levels

at which any information processing system can be understood



- **Computational Theory Level**
  - What are the goals of the computation?
  - What is being computed?
  - Why are these the right things to compute?
  - What overall strategy is followed?

What and Why?

- **Representation and Algorithm Level**
  - How are these things computed?
  - What representation and algorithms are used?

How?

What is the ‘what & why’ of “commonsense world knowledge”

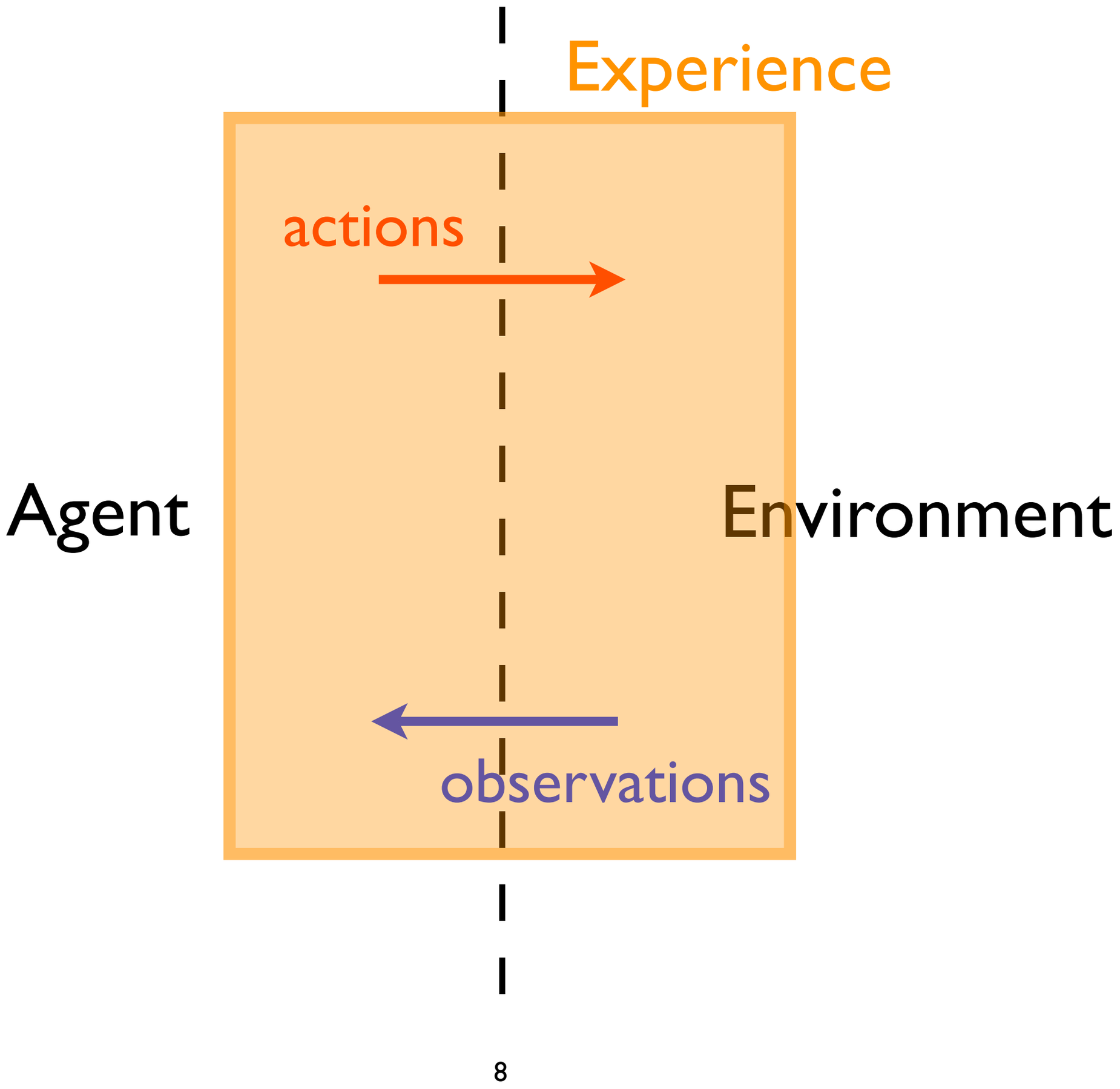
- What computational role does it play?
  - It’s like a world model of course - it’s predictive
  - But some kind of very sophisticated model
    - ✦ a mix of overlapping time scales
    - ✦ interplay of non-intrinsic dynamics and state
- What is the ‘what & why’ of this sophistication?
  - something to do with fast learning and re-planning
- Where should the joints go?

Predictive knowledge

- "John is in the coffee room"
- "My car in is the South parking lot"
- What we know about navigating this hotel/workshop
- What we know about how an object looks, rotates
- What we know about how objects can be used
- Recognition strategies for objects and letters
- "The portrait of Washington on the dollar in the wallet in my other pants in the laundry, has a mustache on it"

Outline

- AI should be oriented around experience
- What do we know?
- Grand challenge of grounding knowledge in experience


Experience
actions
Agent
Environment
observations

Experience is the data of AI

observation

action

$o_1\ a_1\ o_2\ a_2\ o_3\ a_3\ o_4\ a_4\ o_5\ a_5\ o_6\ a_6\ o_7\ a_7 \cdots$

past

now

future

The special thing about life is that it has a now

Experience is what life is all about

Experience is the final common path,
the only result of all that goes on
in the agent and world

Experience matters

- Experience is the most prominent feature of the computational problem we call AI
- It's the central data structure
- It has a definite temporal structure
  - revealed and chosen over time
  - speed of decision is important
  - order is important
- This has unavoidable implications for AI

Outline

- AI should be oriented around experience
- What do we know?
- Grand challenge of grounding knowledge in experience

What do we know? (1)

- We know how to define the problem
  - reward maximization
  - for arbitrary, unspecified reward functions
    - ✦ I consider this to be the very definition of intelligent behavior
  - from low-level experience, say 1000 Hz
- We don't have to change any of this, it is good enough

What do we know? (2)

- We know how to do RL
  - for policies and value functions
- We know how to do planning
  - RL on hypothetical experience, etc
- We understand the interrelations between learning and planning
- We know how to form and use world models flexibly, as in Dyna

What do we know? (3)

- There remain technical issues
  - best way to do function approximation
  - off-policy learning
  - constructing state
- But these *are* details; they don't change the overall outline, the 'what & why' of AI

Summary: the problem of AI

- Observations, actions and rewards at 1000hz
- Maximize reward
- Act faster than you can plan
- Perceive more than you can remember
- There is a now

What are the joints?

- Perhaps they are the boundaries between state representations and their dynamics
  - states and dynamics would have to be co-evolved
- Perhaps we can think of them as a parsing of experience
  - but at multiple, overlapping levels

What makes for good joints?

- Relatively stable dynamics where there is not a joint
- Occasions for changes at the joints
- Compress many steps
- Are the joints the places which would make good states?

Outline

- AI should be oriented around experience
- What do we know?
- Grand challenge of grounding knowledge in experience

Experiential knowledge hypothesis:

All world knowledge is a prediction or memory of sensori-motor experience

- Knowledge is subjective
- Knowledge is ultimately low-level
- Logic and math are not world knowledge
  - they are true in any world

A Grand Challenge:

Grounding knowledge in experience

- To represent human-level world knowledge solely in terms of lowest-level experience
  - sensations
  - actions
  - time
- A minimal ontology
  - no objects, no people, no space, no self, no chickens...
  - all these are "just" patterns in sensation & action

What would it be like to accept the challenge?

- Abstraction is key
  - abstract states (eg, predictive representations)
  - abstract actions/transitions (eg, options)
- Need to think in unfamiliar ways
- Microworlds, robotics
- Indexical (deictic) representations
  - sequence instead of symbols

In experiential terms,

- What is space?
  - regularities in sensation change with eye movement
- What are objects?
  - subsets of sensations
  - that tend to occur together temporally
  - and can be in arbitrary relative spatial arrangements

- What is my body, my hands?
  - objects that are always present
  - and can be controlled
- What are people?
  - objects that may move on their own
  - that have a particular subset of sensations
  - whose presence may change my sensations for the better
  - eventually:
    - ✦ that are best predicted with respect to goals
    - ✦ that are analogous to me

Relational ⇒ Indexical

∀ objects X, If I drop X, then X will be on the floor

- Holding object X means predicting certain sensations if, for example, one directs one's eyes toward one's hand
- Thus, on dropping, the predicted sensations are merely transferred from the looking-at-hand prediction to the looking-at-floor prediction
- Such transfer of existing predictions should be a common part of visual knowledge - updated every time the eyes move

∃ X,Y, such that Red(X), Blue(Y), and Above(X,Y)

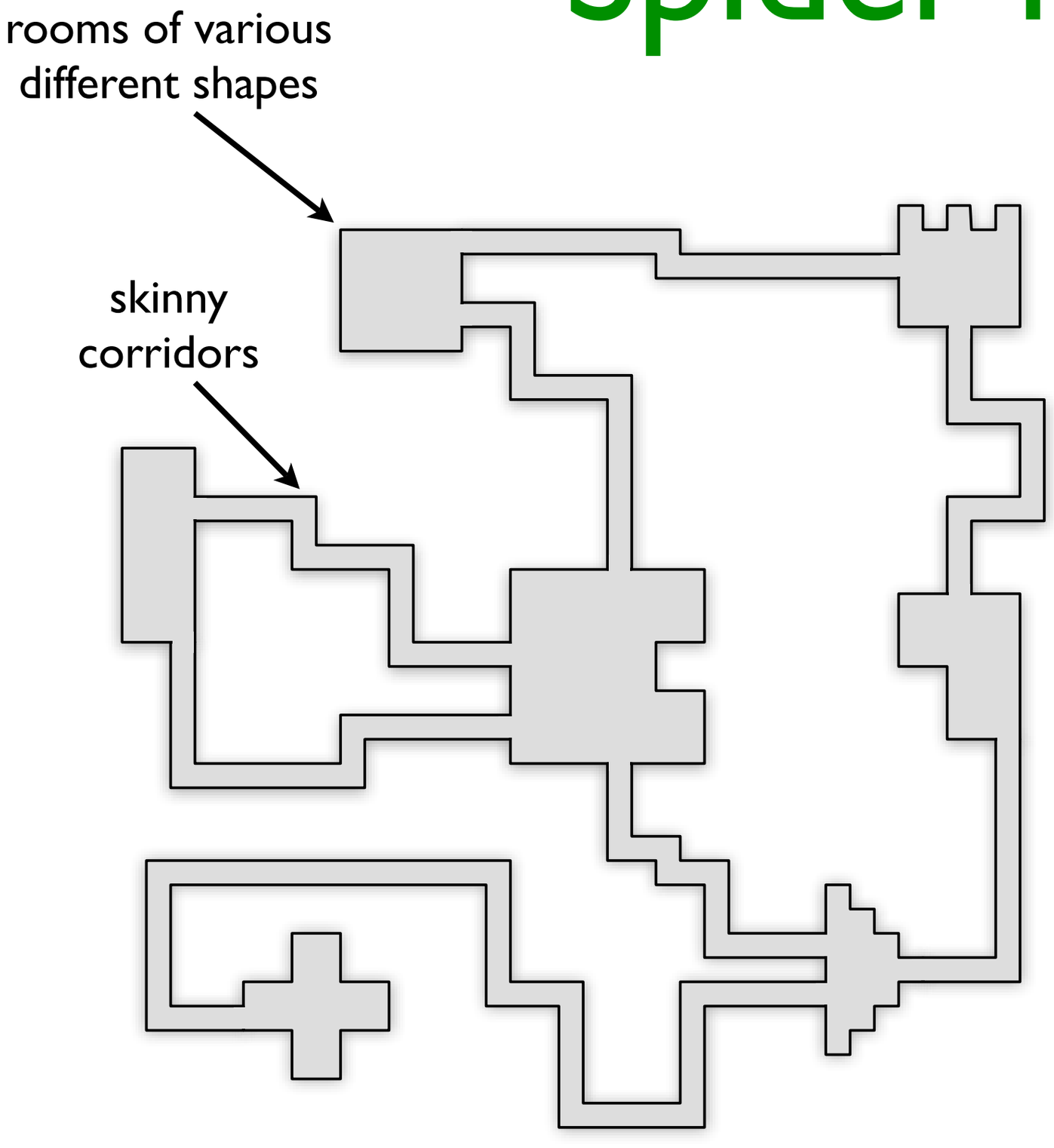
- There is some place I can foveate and see Red
- There is some place I can foveate and see Blue
- If I foveate first the Red place, "mark" it, then the Blue place, the mark will be *above* the fovea (repeat until succeeds)

These are typical ideas of modern, active, deictic vision

Conclusion:
Research strategies re world knowledge

- It's all about carving the world at its joints
  - so that we can adapt to change/data faster
- But we don't want to find the joints ourselves
- Nor are we ready to automate it
- We need the language/framework within which to do it
- And the intuition to imagine it

Spider maze



- Single-bit experience
  - obs: wall/open
  - action: forward/turn
- Rooms are a pattern of interaction - PSR
- Options for following corridors
- Can we represent, learn, discover, and plan with the relevant knowledge?

Thank you for your attention