

How do we understand an intelligent system?

Readings for today

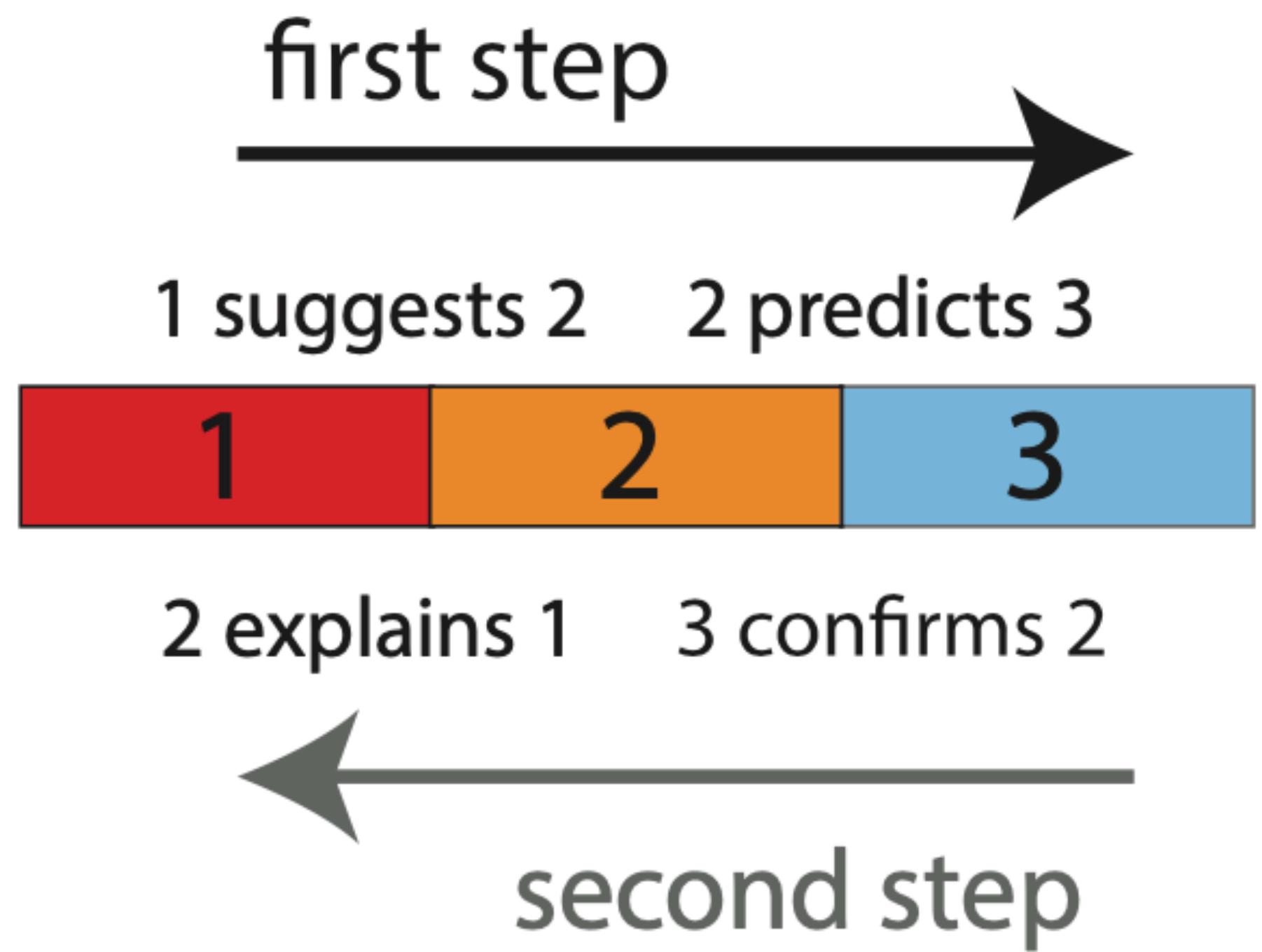
- Braitenberg, V. (1986). Vehicles: Experiments in synthetic psychology. MIT press.

Topics

- The limitations of empirical approaches
- The synthetic psychology approach

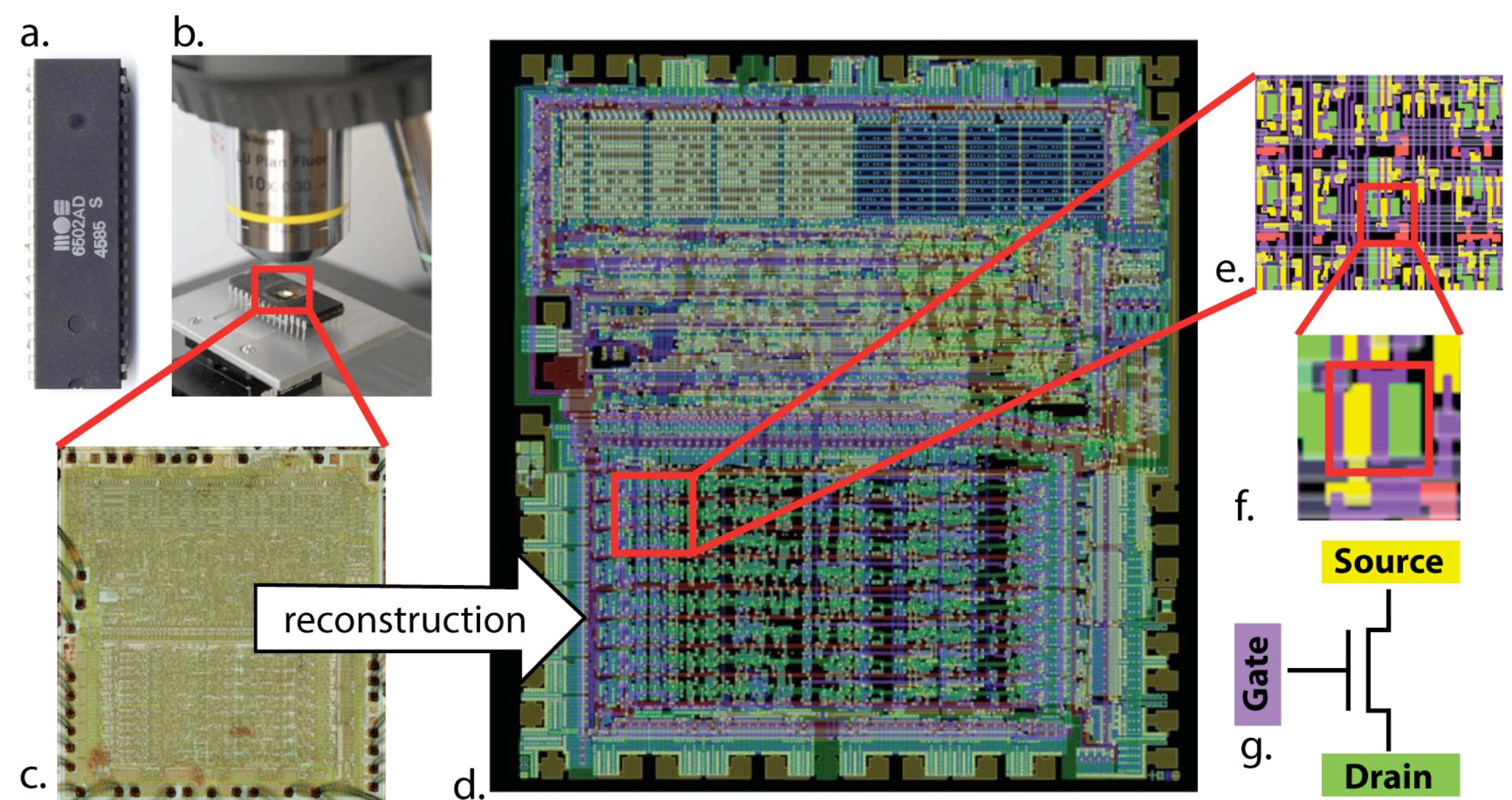
The limitations of empirical approaches

Marr's levels of analysis

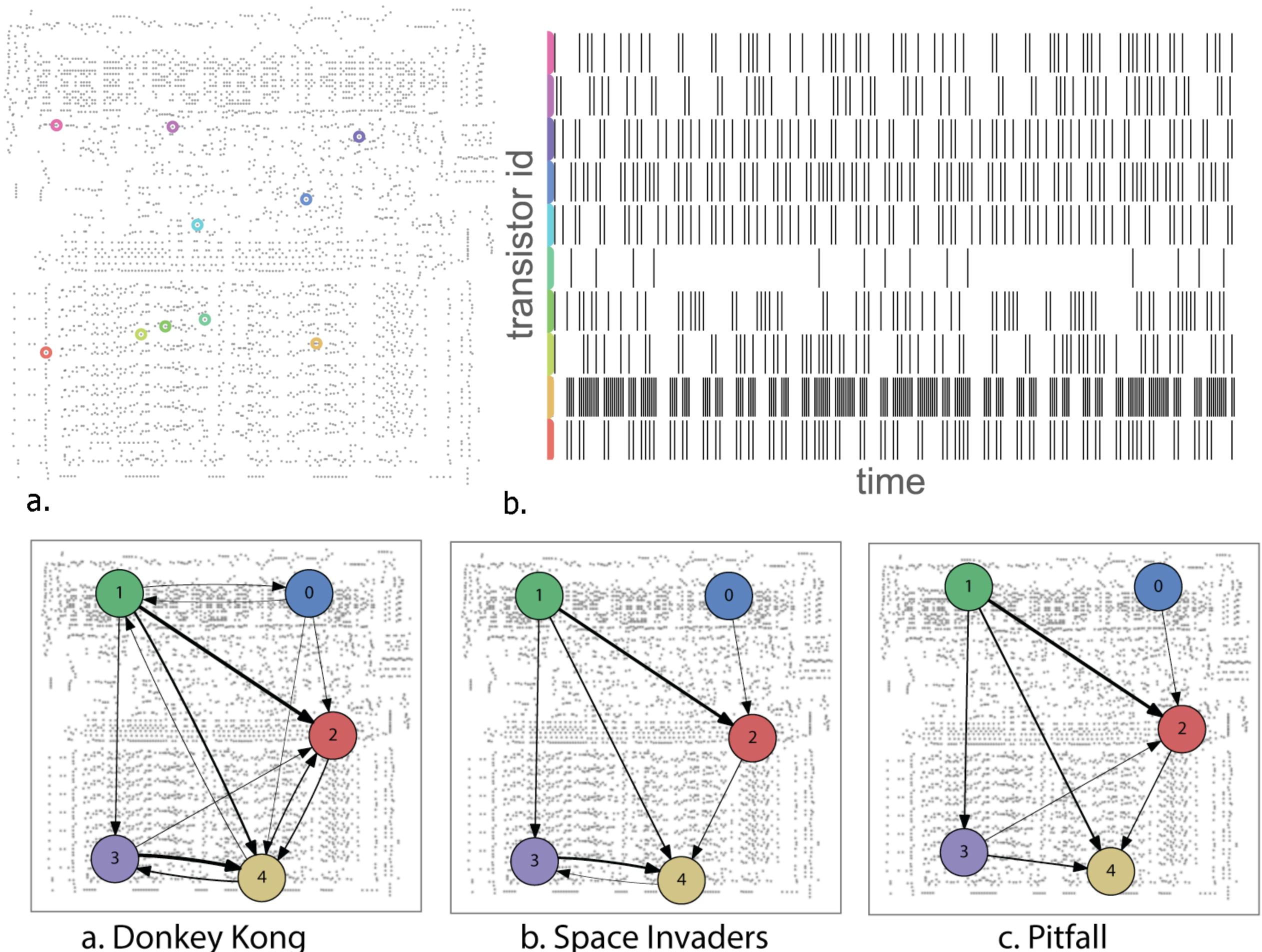


Could we understand a microprocessor?

Microprocessor

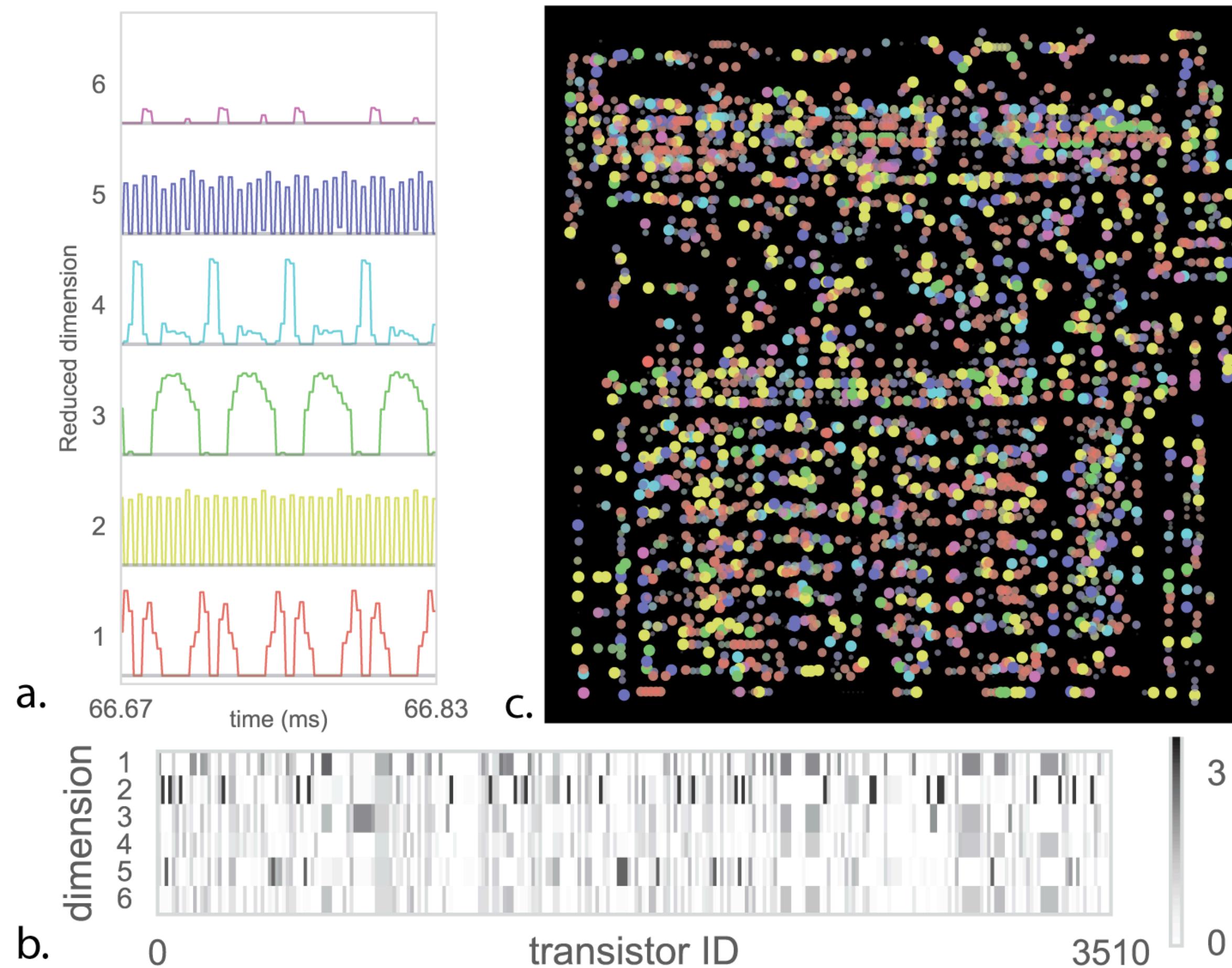


Neuro techniques



Could we understand a microprocessor?

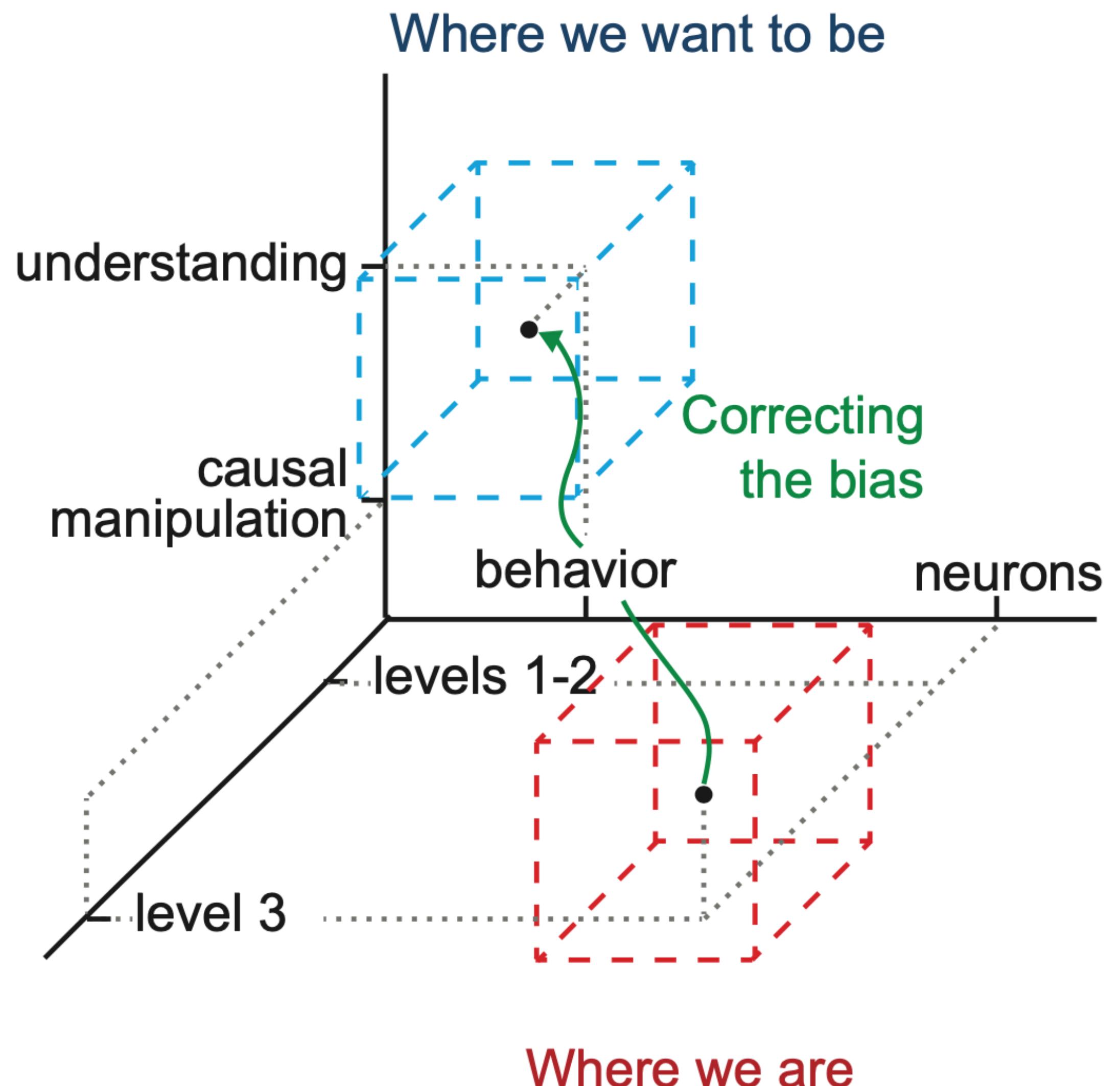
Inferred system



Conclusions

- The microprocessor has a clock mechanism of some kind.
- The microprocessor has a low-dimensional state.
- Unable to reconstruct other known properties of the microprocessor or the chip itself.

Where we need to be



“Understanding something is not the same as just describing it or knowing how to intervene to change it.”



Theory

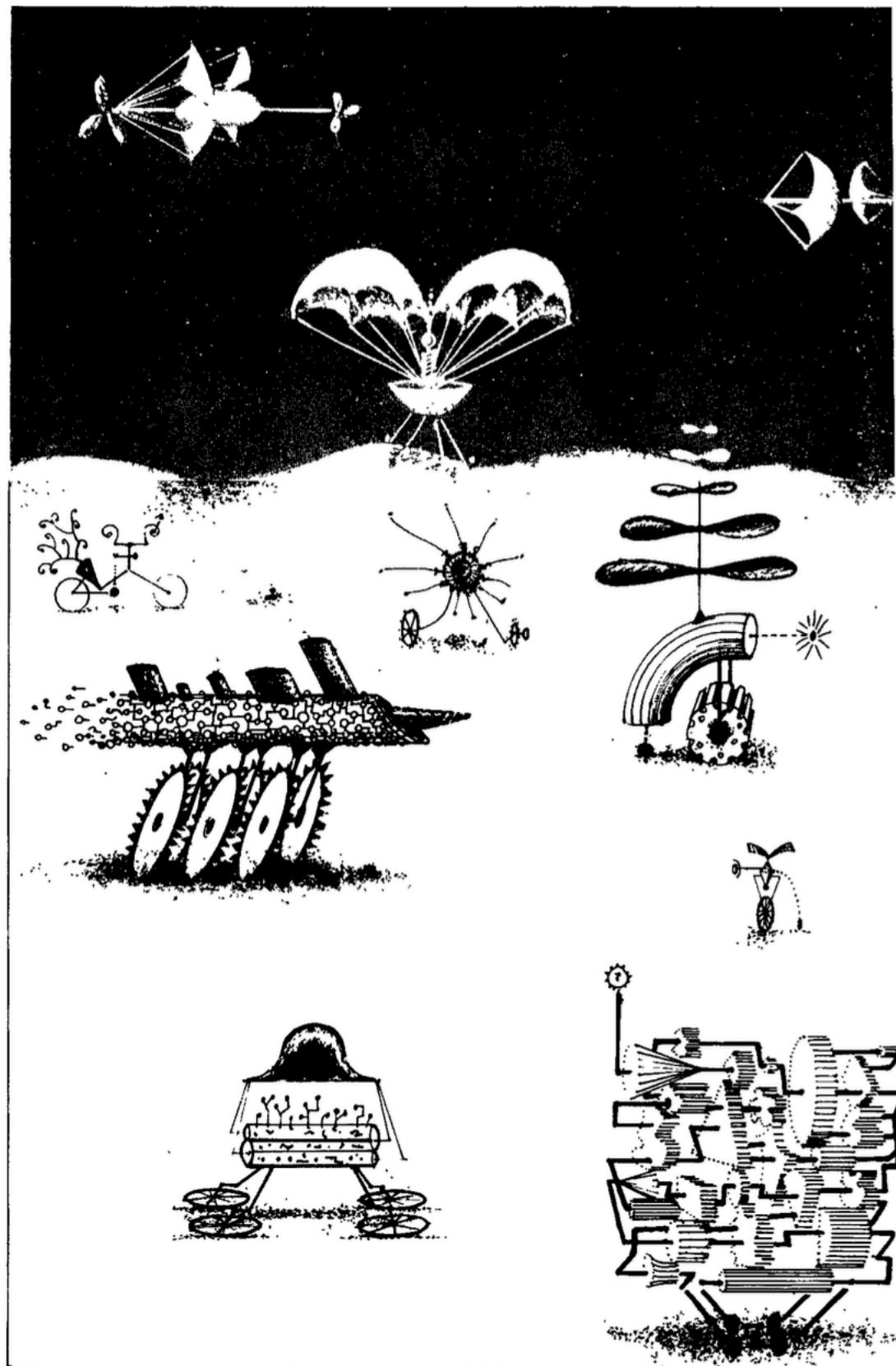
The synthetic psychology approach

The concept of Vehicles

Basic idea: Imaginary vehicles with simple mechanical systems that exhibit complex behaviors.

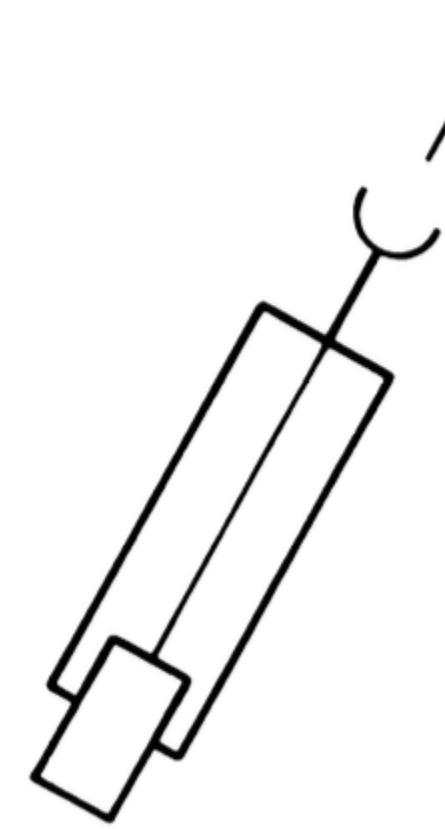
Goal: Understand how simple, neural-like circuits can produce complex actions.

Insight: Complex behavior can emerge from simple rules and structures.

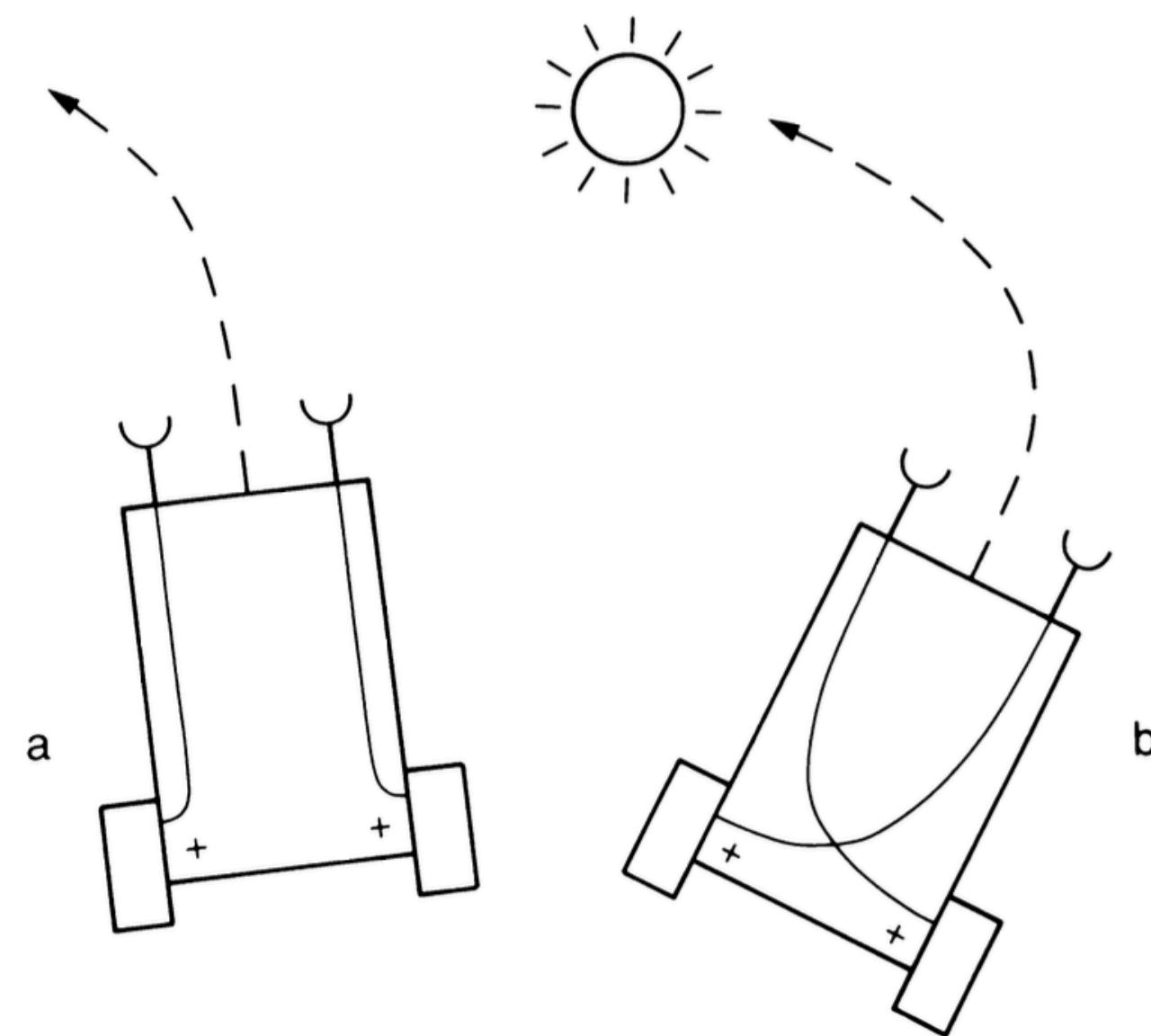


Vehicles 1-3

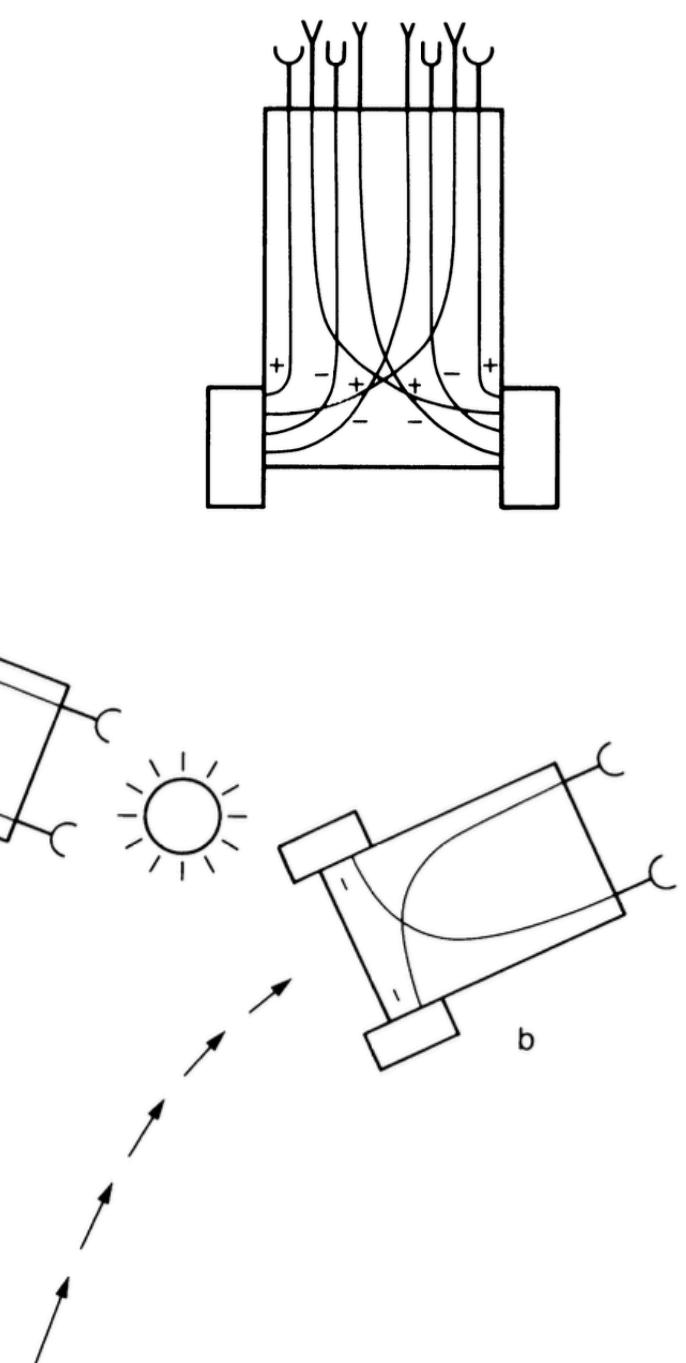
Simple reflexive behaviors such as moving towards light or heat.



Vehicle 1



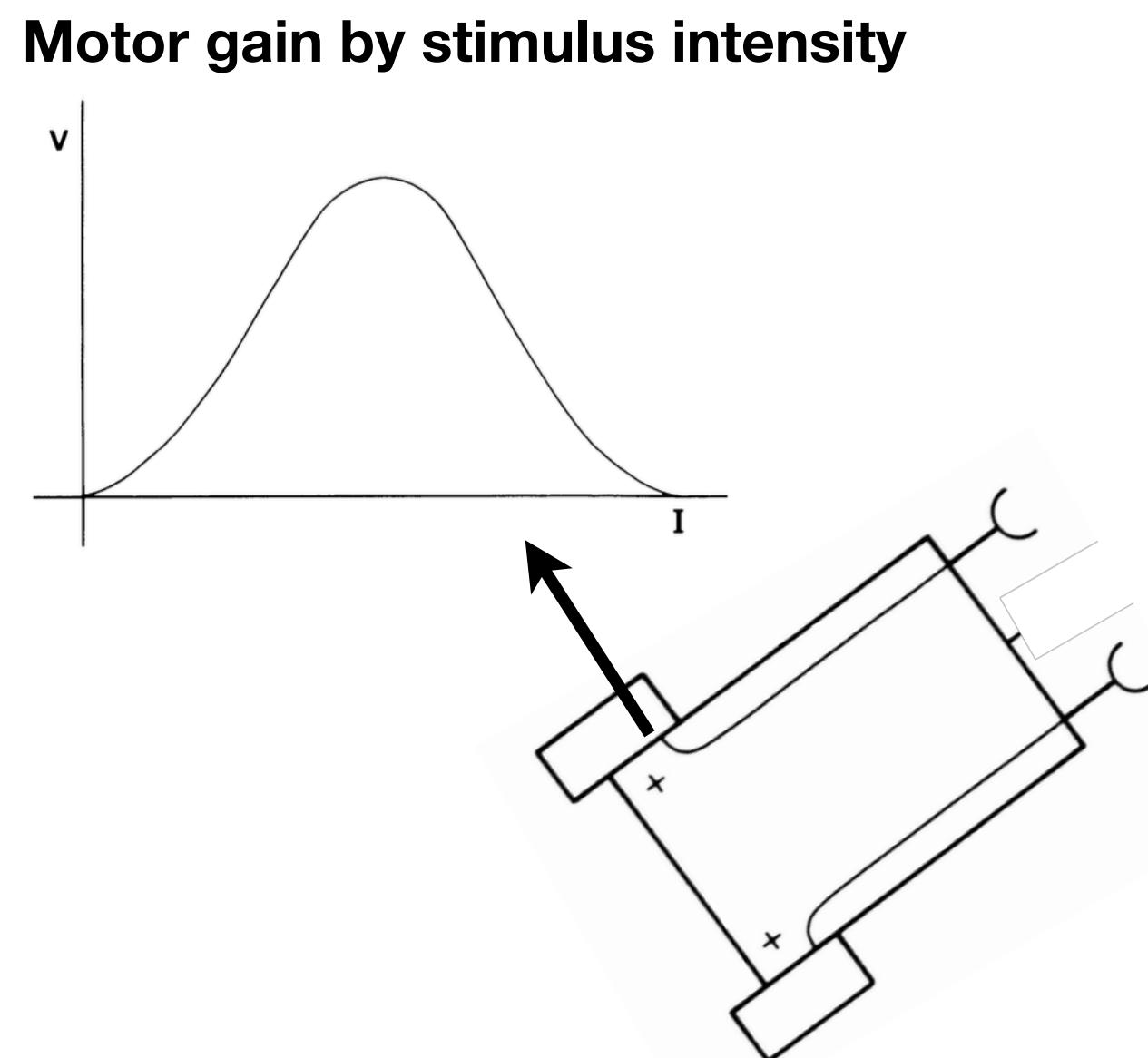
Vehicles 2a-b



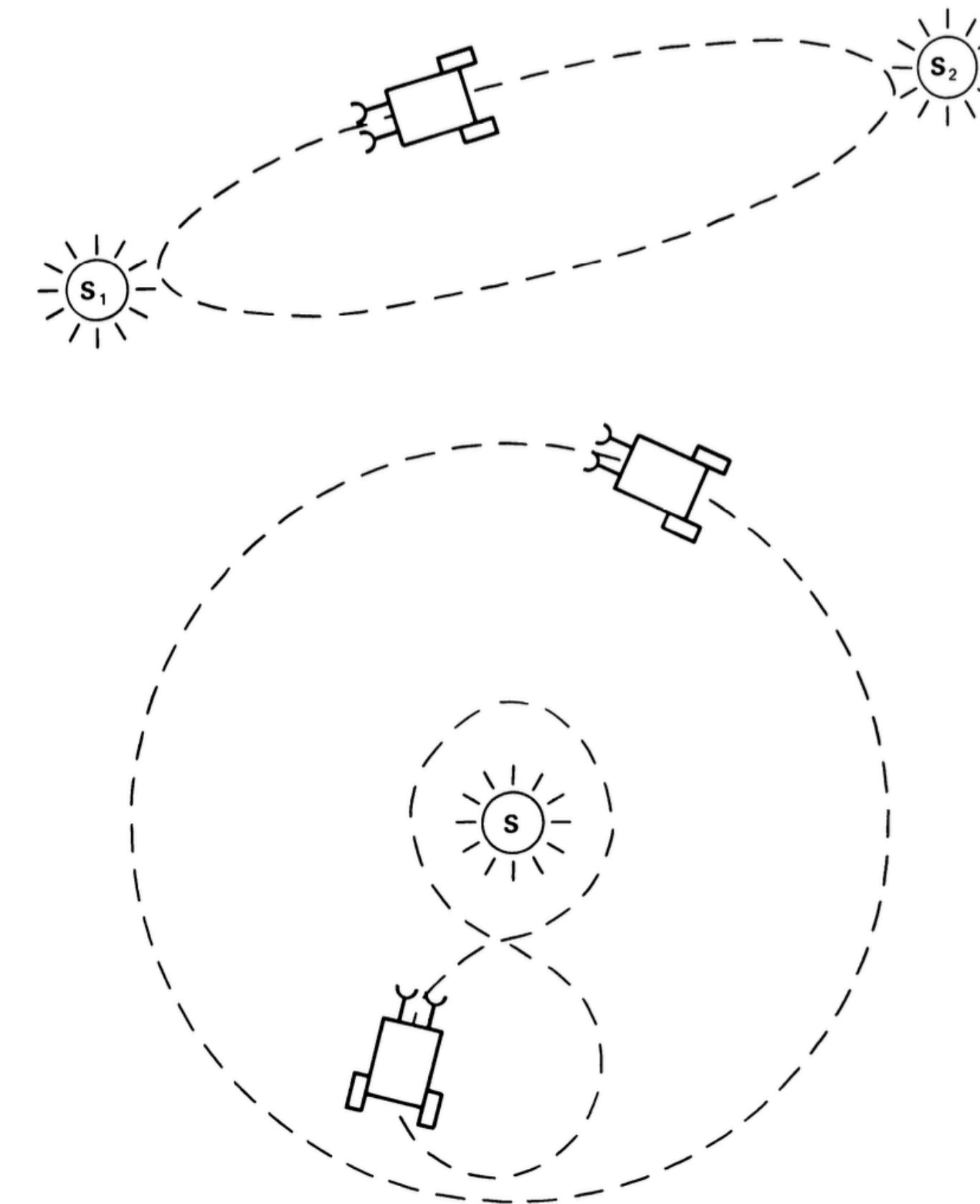
Vehicles 3

Vehicles 4-7

More complex behavioral patterns arising from more complex sensors and actuators.



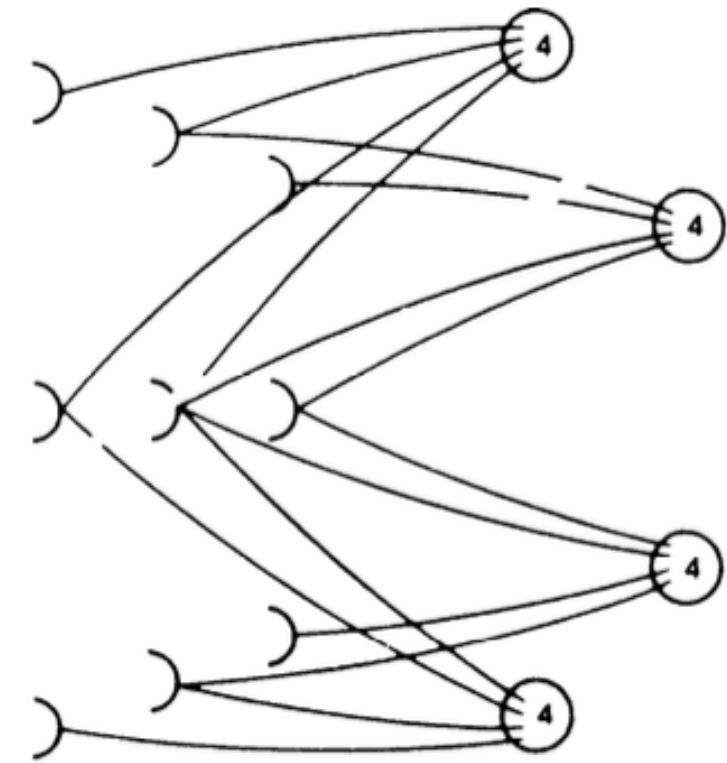
Vehicle 4



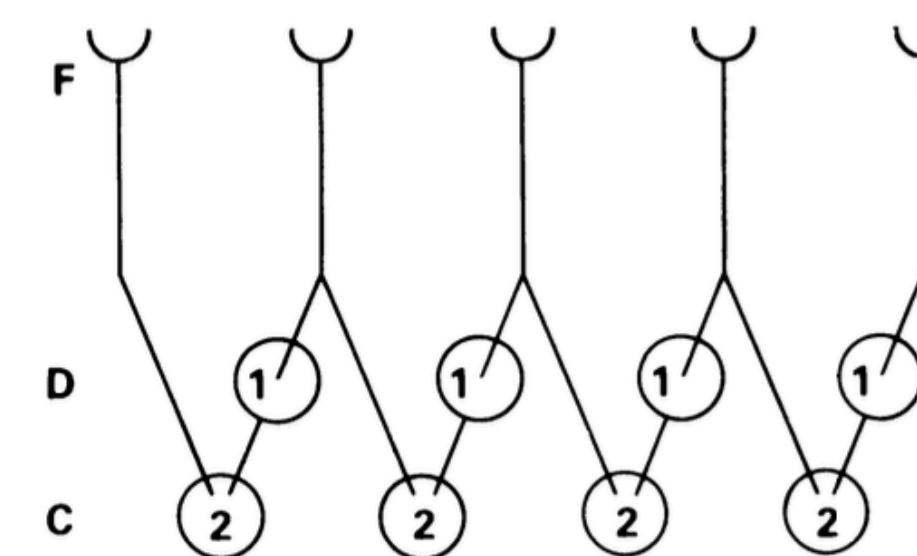
Vehicles 8-14

Advanced internal representations and “cognitive” functions

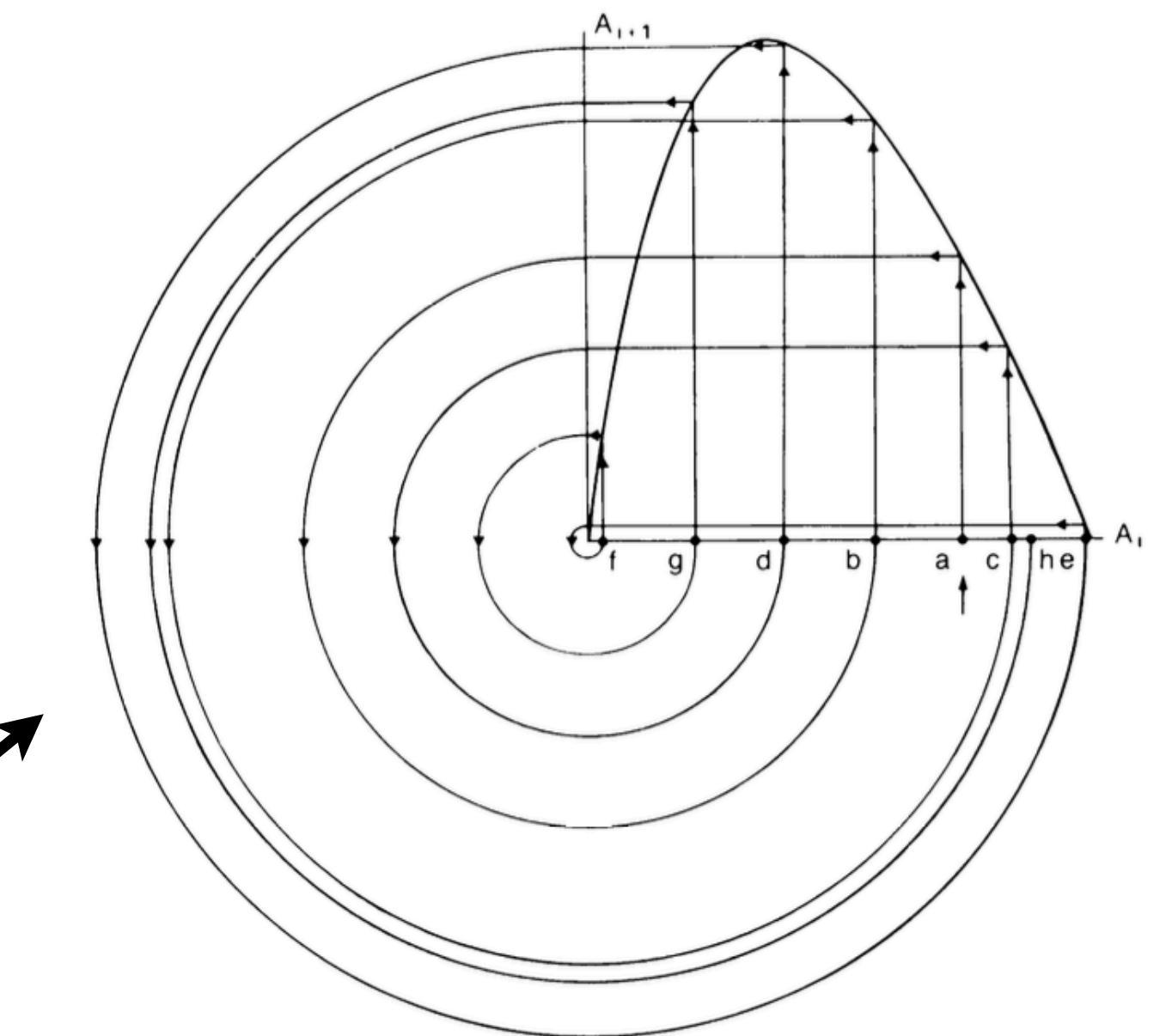
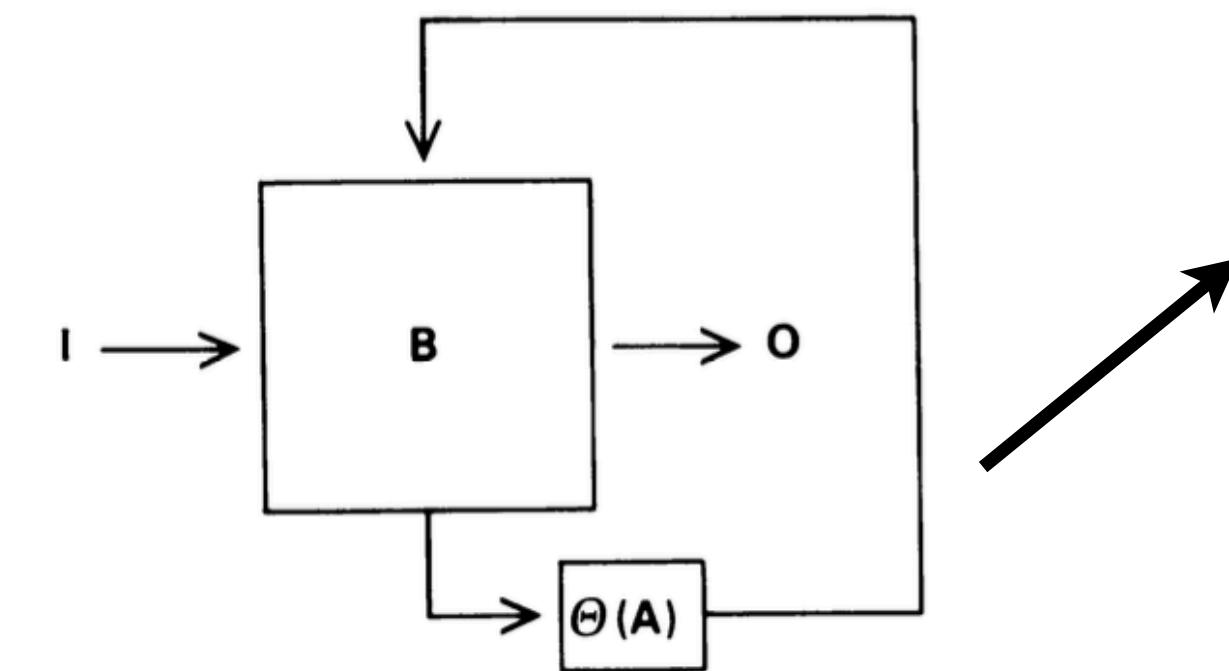
Object detection architecture



Movement detection architecture



Working memory architecture



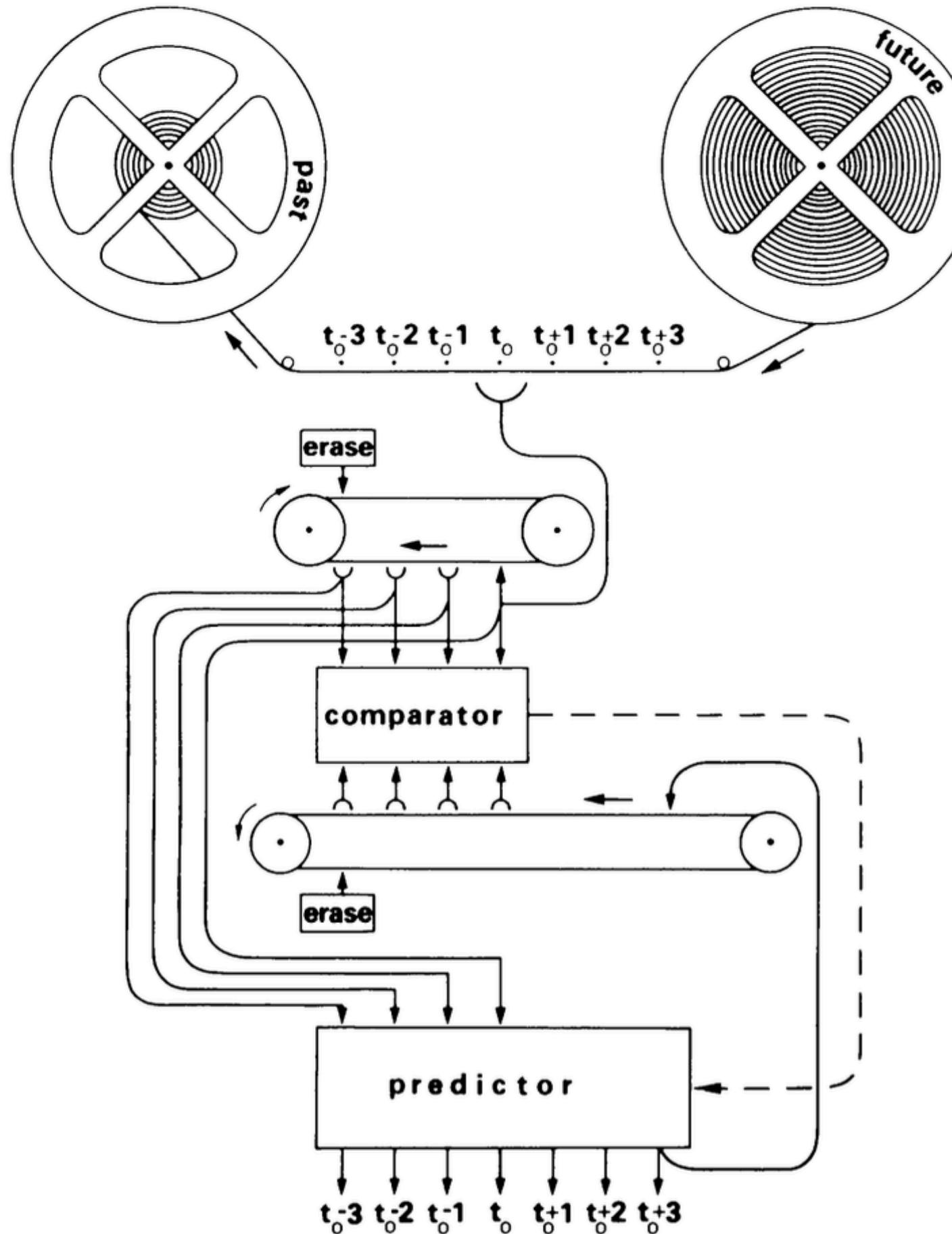
Vehicle 8

Vehicle 12

Vehicles 8-14

Advanced internal representations and “cognitive” functions

Vehicle 13

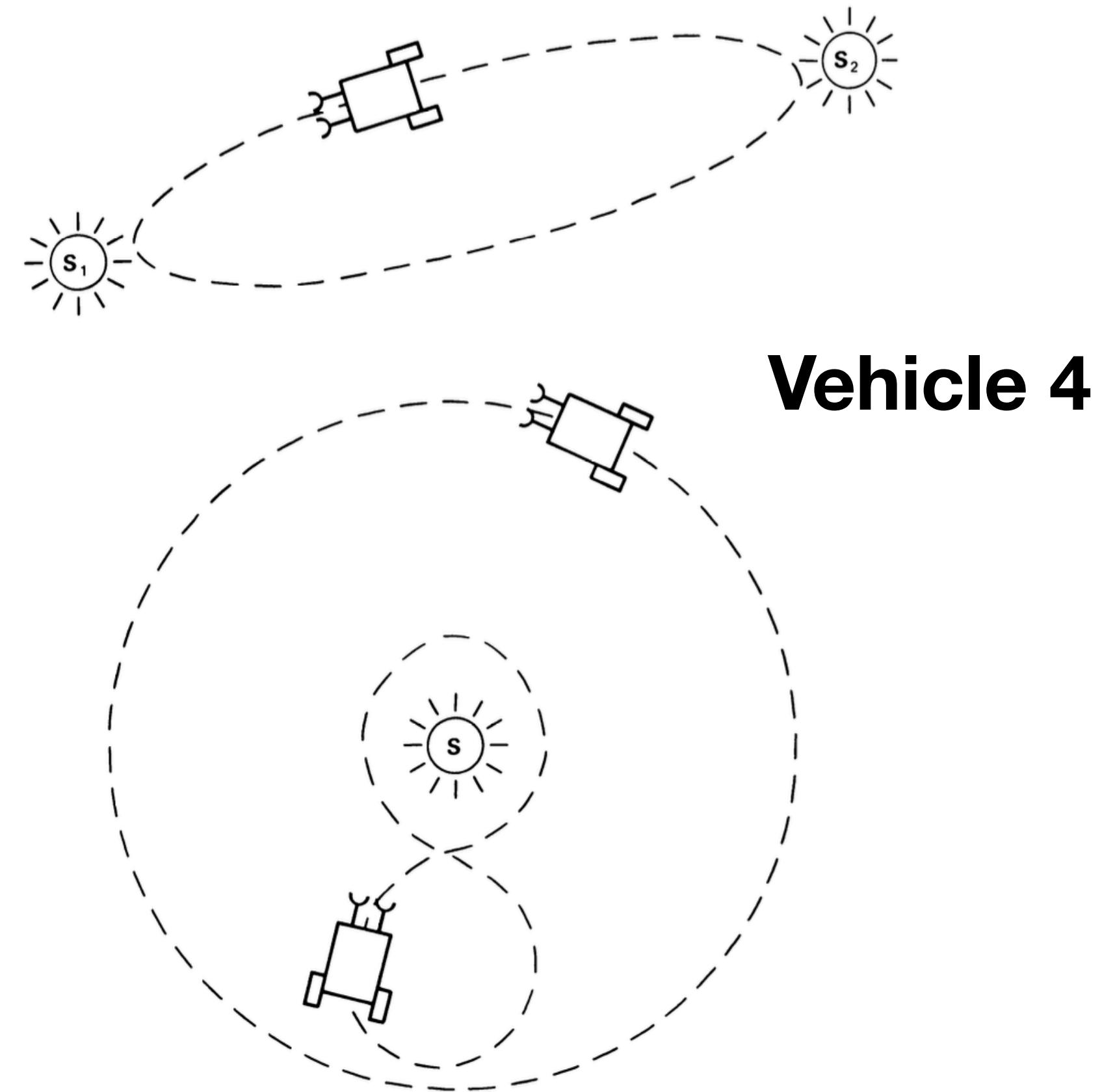


Mechanical system that implements a basic short-term prediction like a Kalman Filter

Emergent Behavior

Definition: The appearance of complex dynamics from simple systems

Implication: Challenges traditional views of behavior and suggests a bottom-up approach to understanding intelligent systems



Take home message

- We need to understand intelligent behaviors at multiple levels: from computation to neurons and back again.
- Current bottom-up approaches (neurons→behavior) are fundamentally limited in what we can understand.
- Synthetic psychology shows how reasoning from simplistic models of implementation mechanisms can offer deep insights into the emergence of agentive behavior

Break out group discussions

What are critical limitations of the synthetic psychology approach that we need to consider when moving forward? Identify both the limitations and what problems they can cause.