Robot Epistemology:
The Problem of Knowledge and Data

Anna Koop

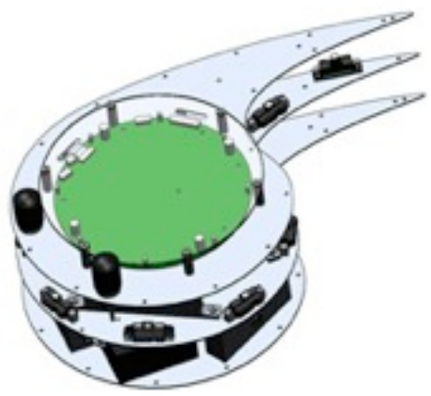
rlai.net

NSERC CRSNG

I

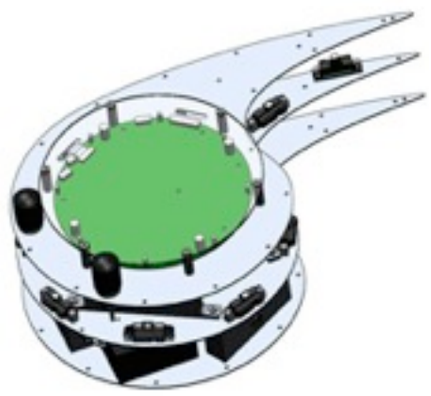
Epistemology: study of knowledge

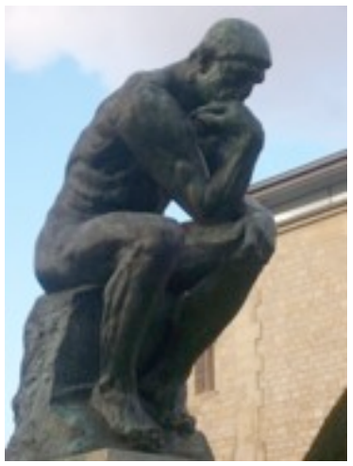
Robot Epistemology



We want to build things
that know things.

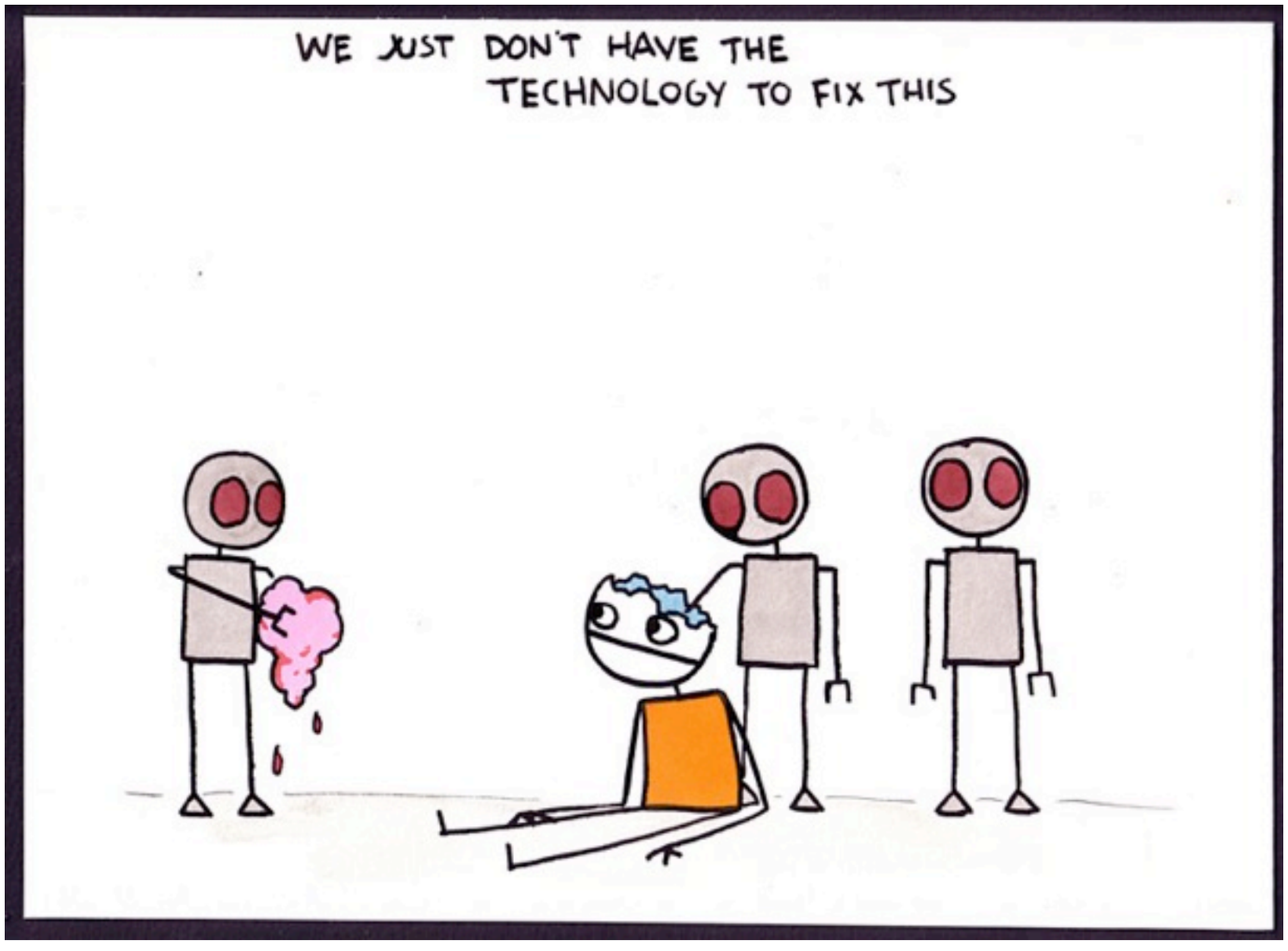
Robot Epistemology





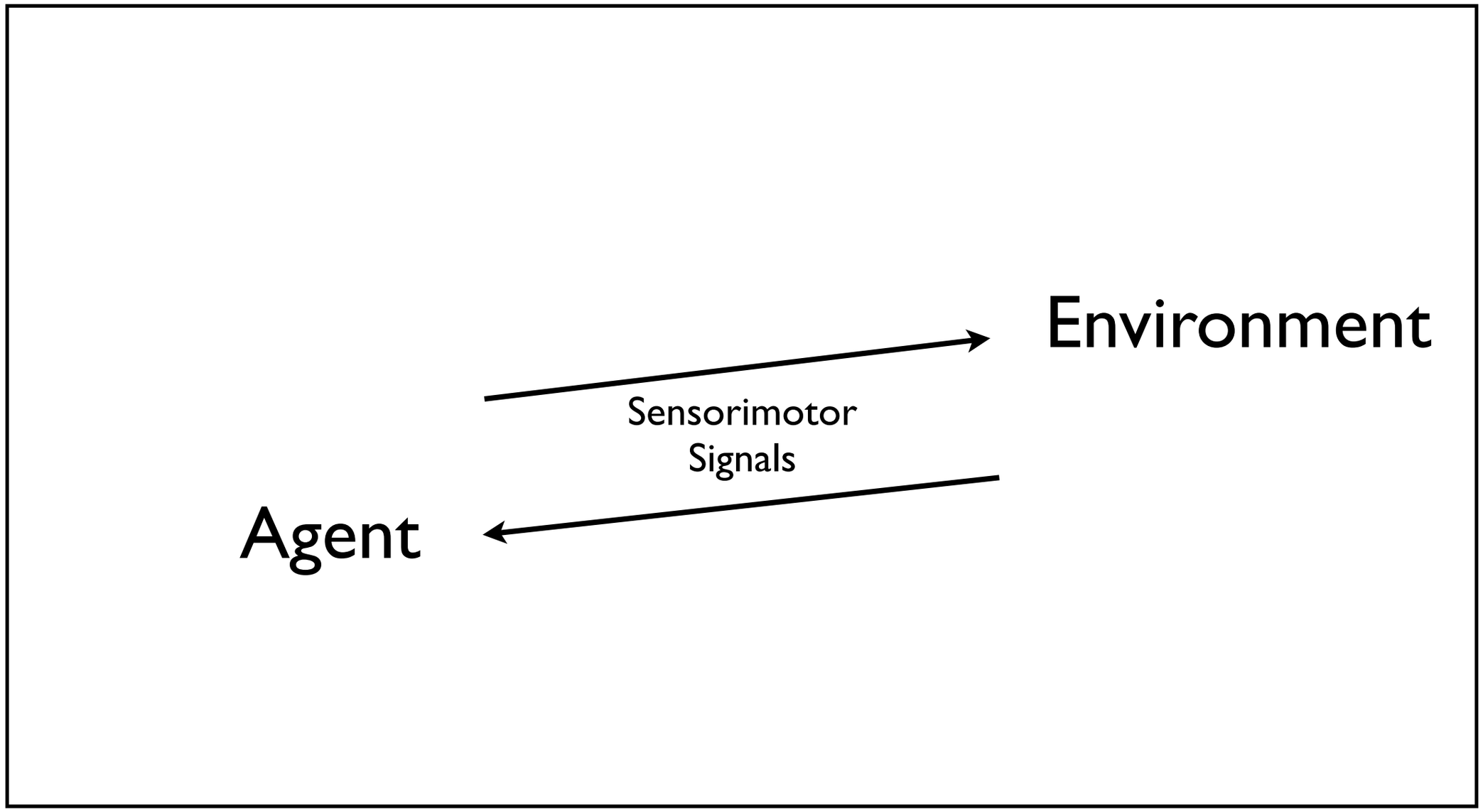
We want to build things
that know things.

Robot Epistemology



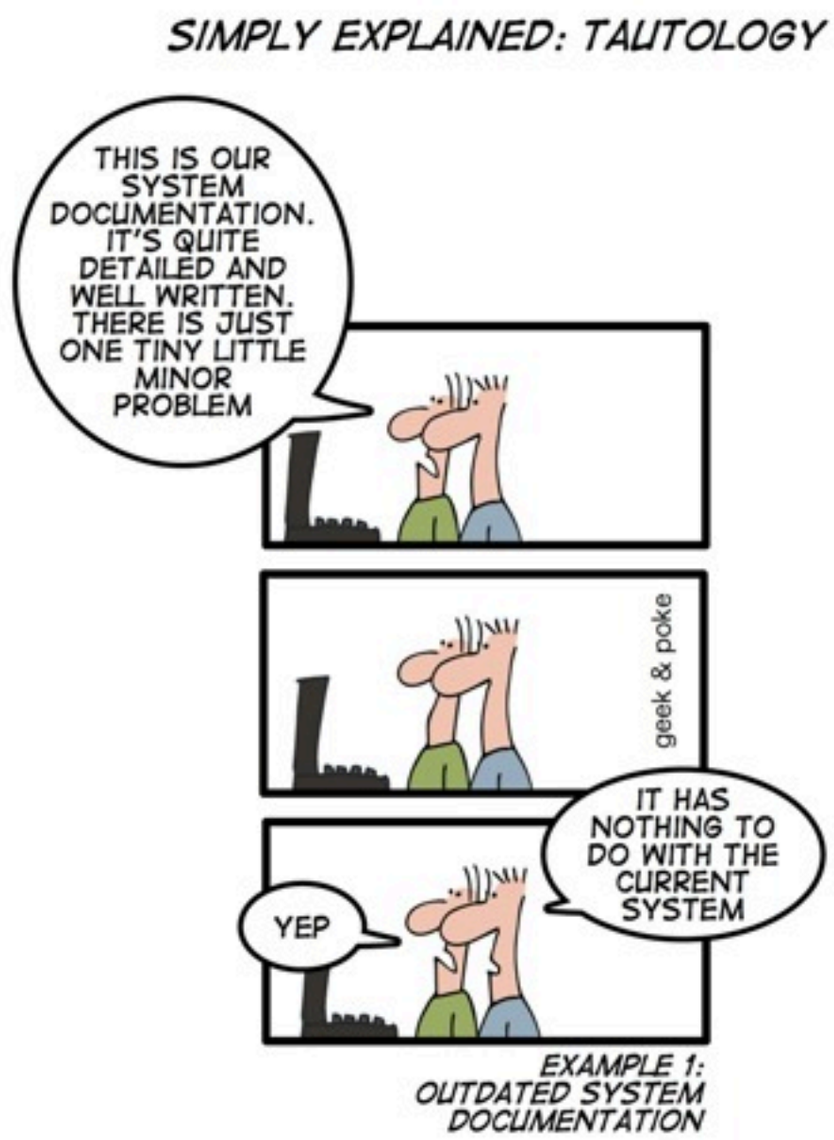
This turns out to be hard. Our attempts at game playing, autonomous driving, question-answering Impressive but falling short of dreams
Why?

Building things that know things



Look at our usual diagram. Drawing the distinction between what's available theoretically, as a researcher, and to the mind itself.

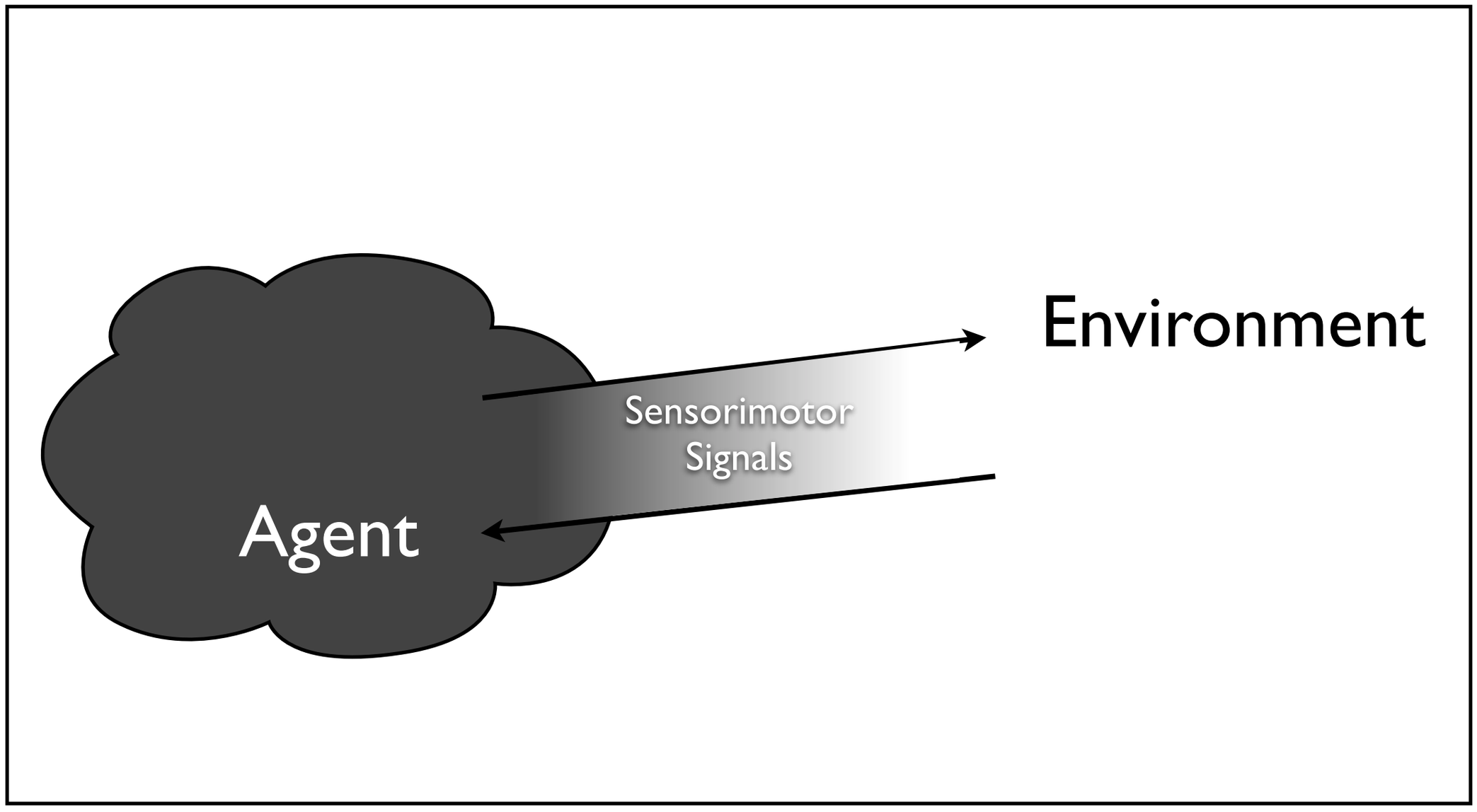
Verification/Validation



There's the question of whether it's doing what we want it to do. Whether the knowledge is appropriate

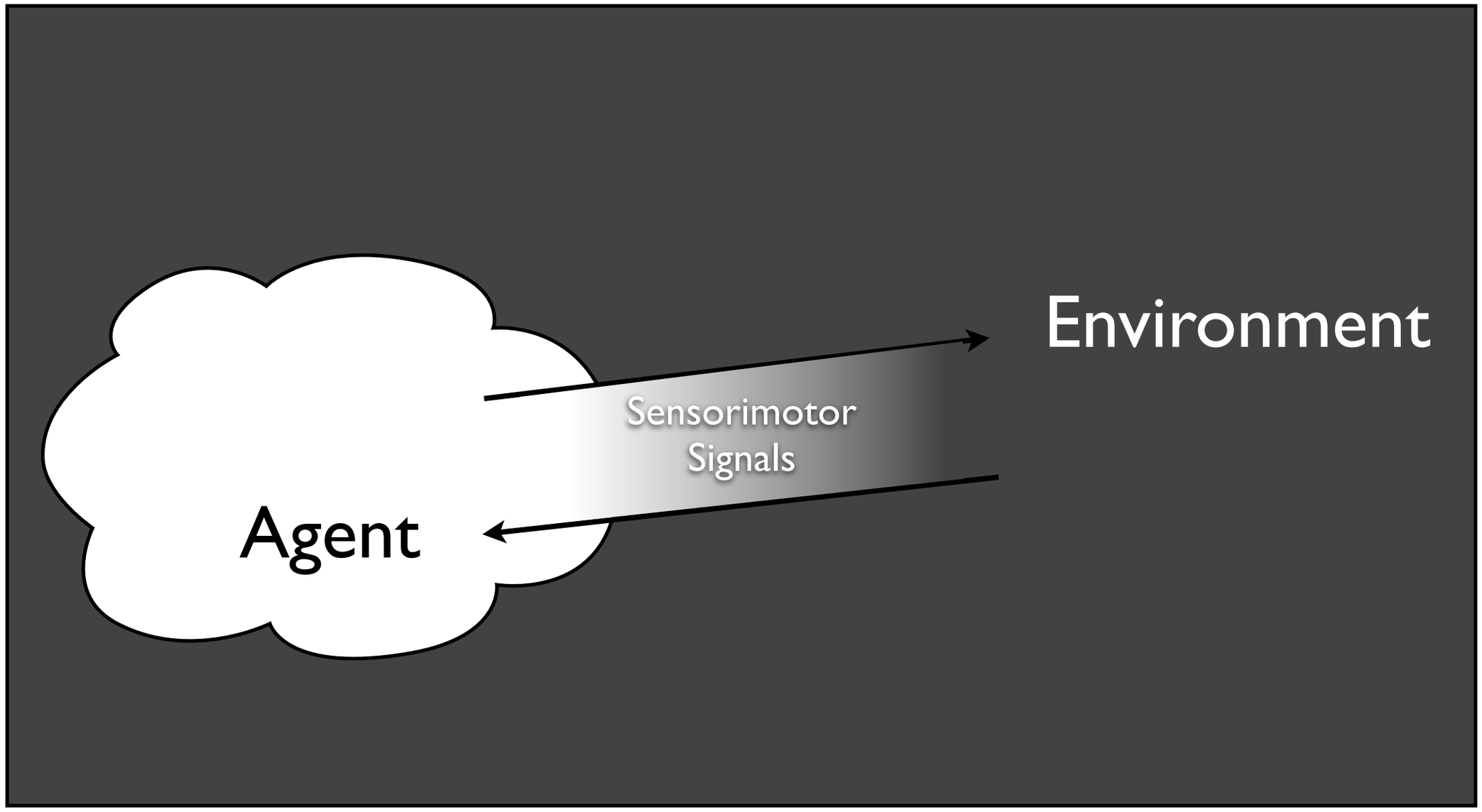
And there's the question of whether its knowledge is accurate. And useful. And consistent. And matches data.

Validation



Is it doing what we want?
Is it appropriate and useful knowledge that looks to us like “knowing things”

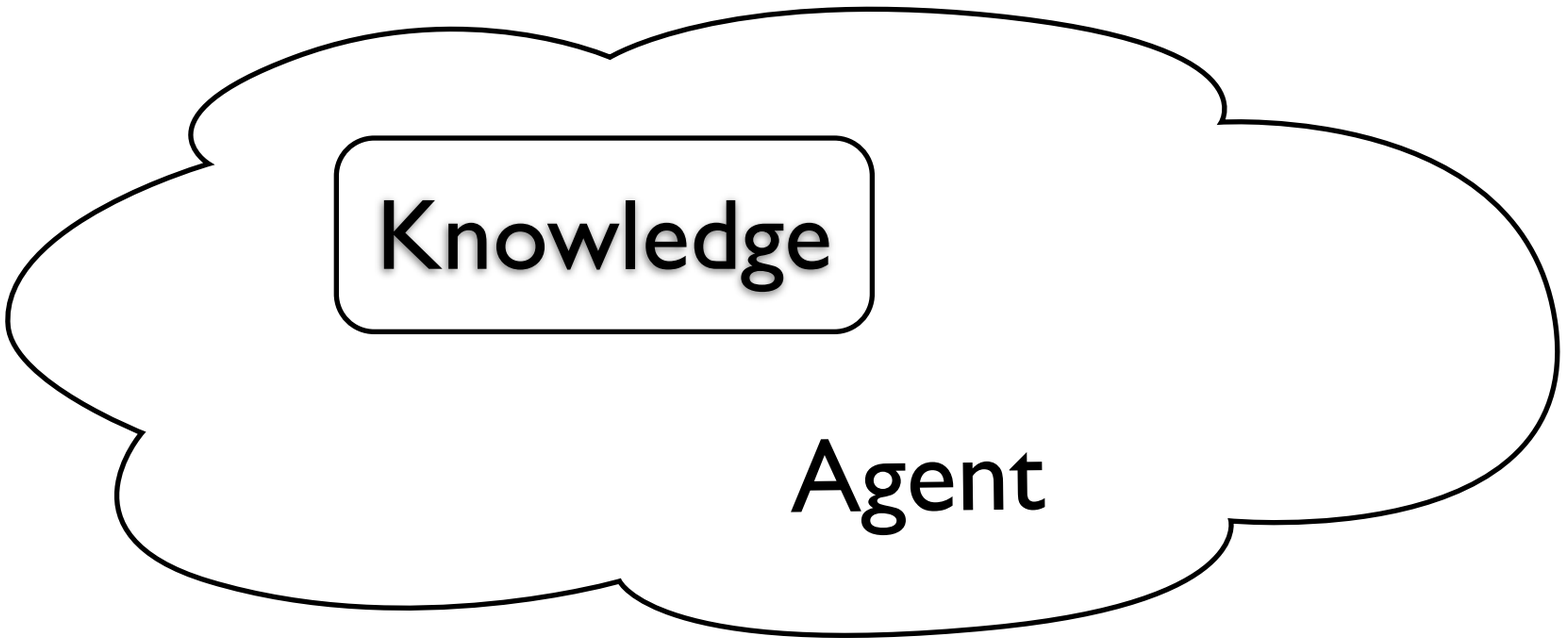
Verification



Engineering internal view ("Are you building it right?")

Robot epistemology: Is what it knows accurate?
"How can the mind itself know that its information is correct?" Lighting up the mind

What it knows



- summary information used by the mind of an intelligent agent

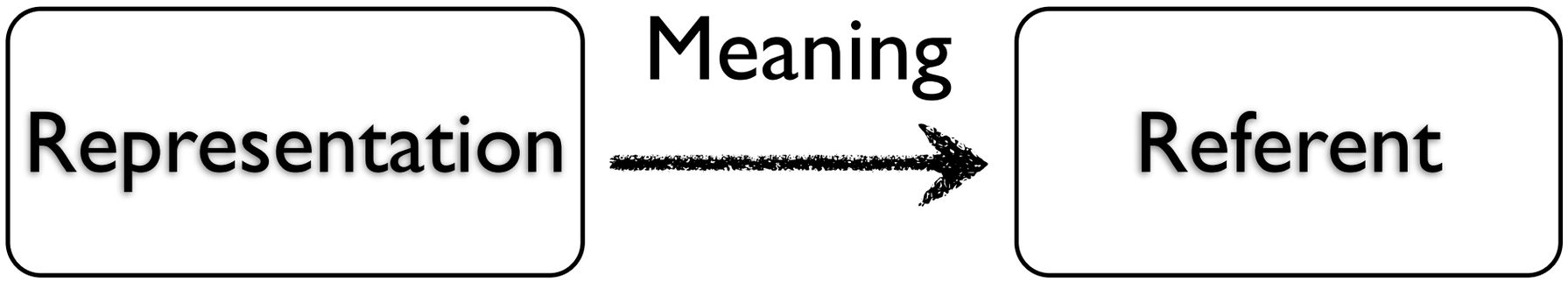
To get the issues of verification clearer we need to talk about knowledge
My definition

Deconstructing Knowledge

- What is knowledge about?

-signifier/signified
- what knowledge is about (key epistemological question)
Key for verification---knowing something is true involves the mapping between the knowledge and its referent

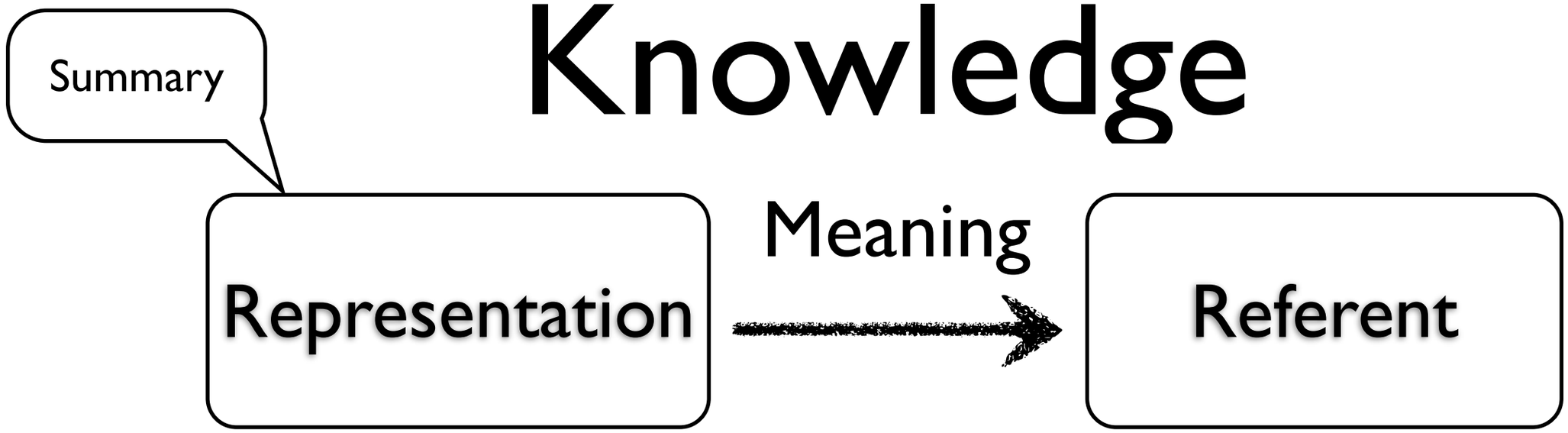
Deconstructing Knowledge



- What is knowledge about?

-signifier/signified
- what knowledge is about (key epistemological question)
Key for verification---knowing something is true involves the mapping between the knowledge and its referent

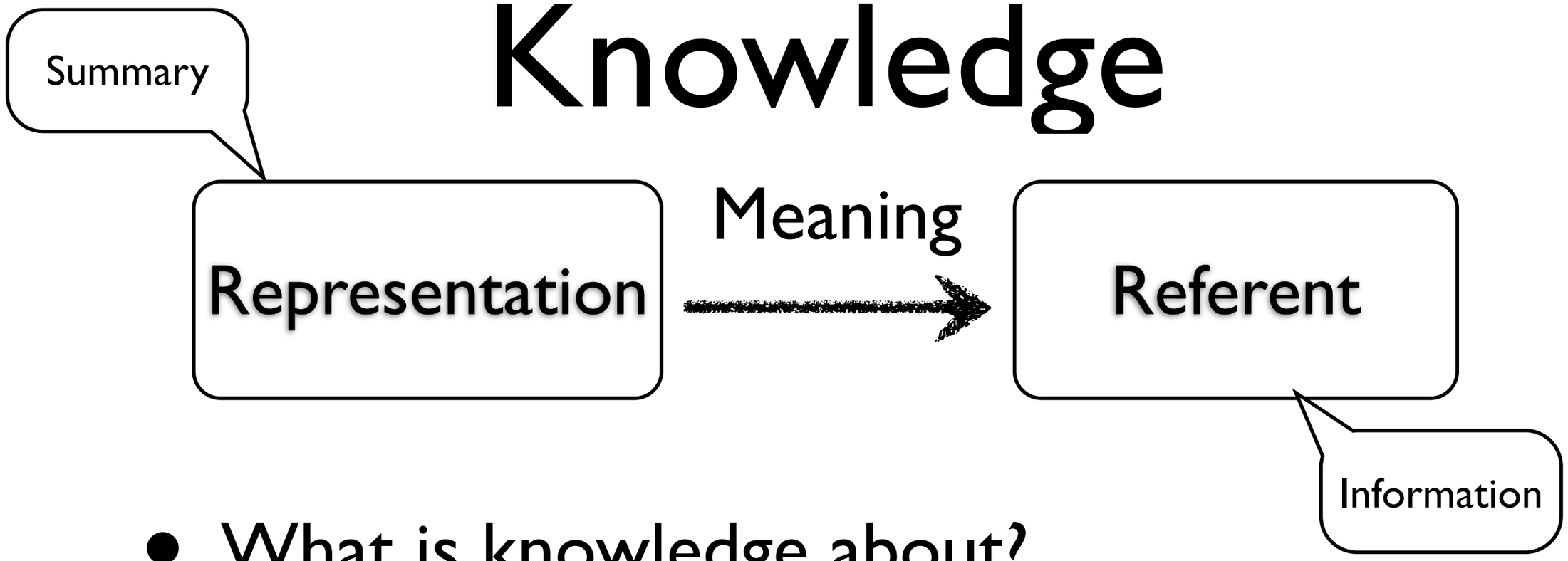
Deconstructing Knowledge



- What is knowledge about?

-signifier/signified
- what knowledge is about (key epistemological question)
Key for verification---knowing something is true involves the mapping between the knowledge and its referent

Deconstructing Knowledge



- What is knowledge about?

-signifier/signified
- what knowledge is about (key epistemological question)
Key for verification---knowing something is true involves the mapping between the knowledge and its referent

Verification

- Check for consistency
- Ask an oracle
- Check against data

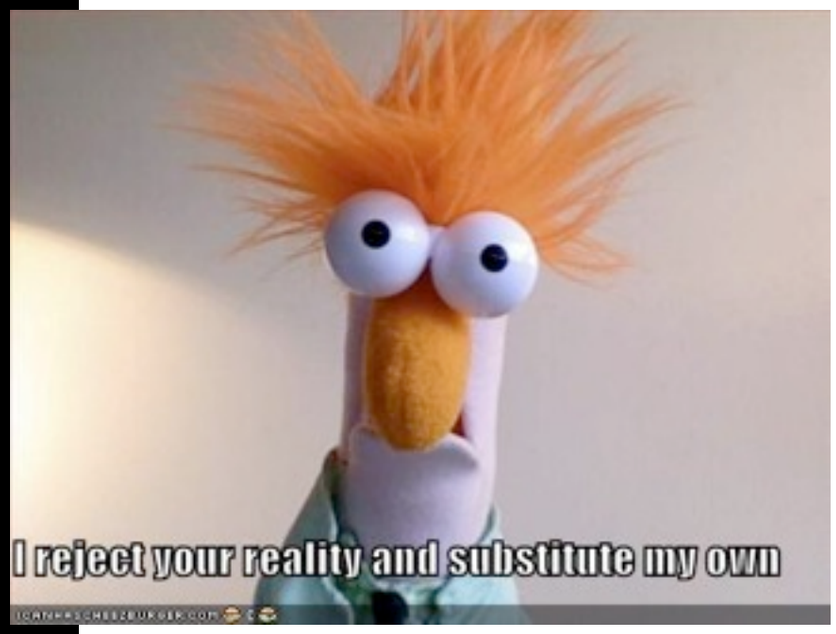
Three methods of verifying, evaluating the truth of the representation

Is it consistent?

$$\neg(P \wedge Q) \equiv \neg P \vee \neg Q$$
$$\neg(P \vee Q) \equiv \neg P \wedge \neg Q$$
$$\neg(\neg P) \equiv P$$
$$\neg(\forall x) P(x) \equiv (\exists x) \neg P(x)$$
$$\neg(\exists x) P(x) \equiv (\forall x) \neg P(x)$$

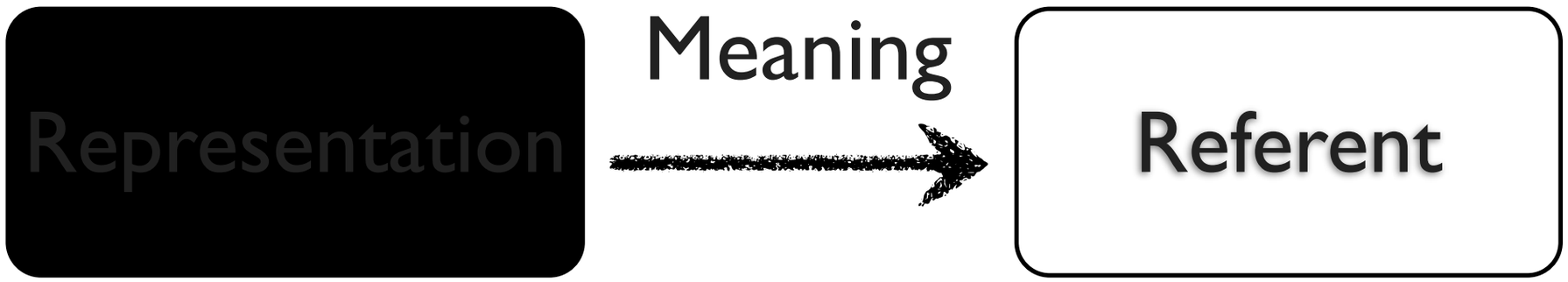
Whether the KB contains contradictions, whether the predictions or belief distribution sums to one, whether the projection ends up in a valid state.
...but also

Is it true?



To evaluate if it's true, need some connection to the referent

Verification



- Check for consistency
- Ask an oracle
- Check against data

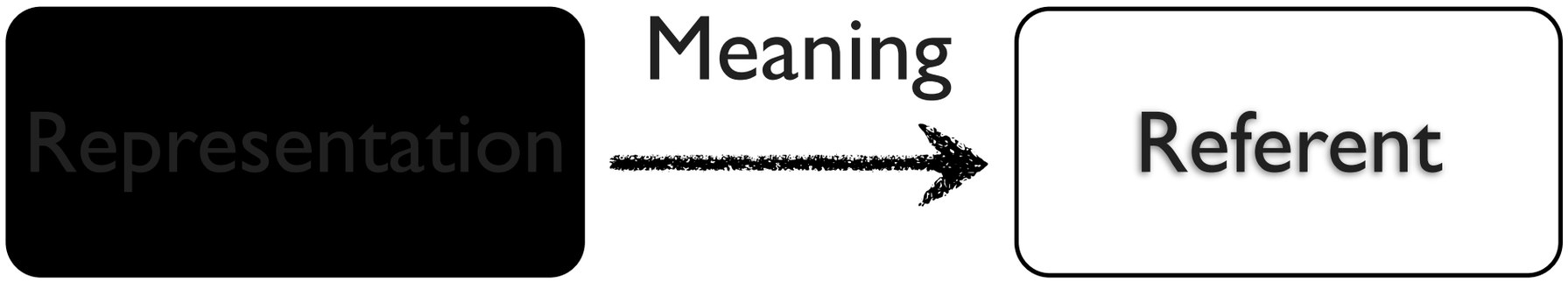
Oracular verification

looking at encoding the stuff that someone has written down in a biology textbook, in a form that can be computed with.

A less extreme example is learning from supervised labels.

Or learning the value function or state representation.

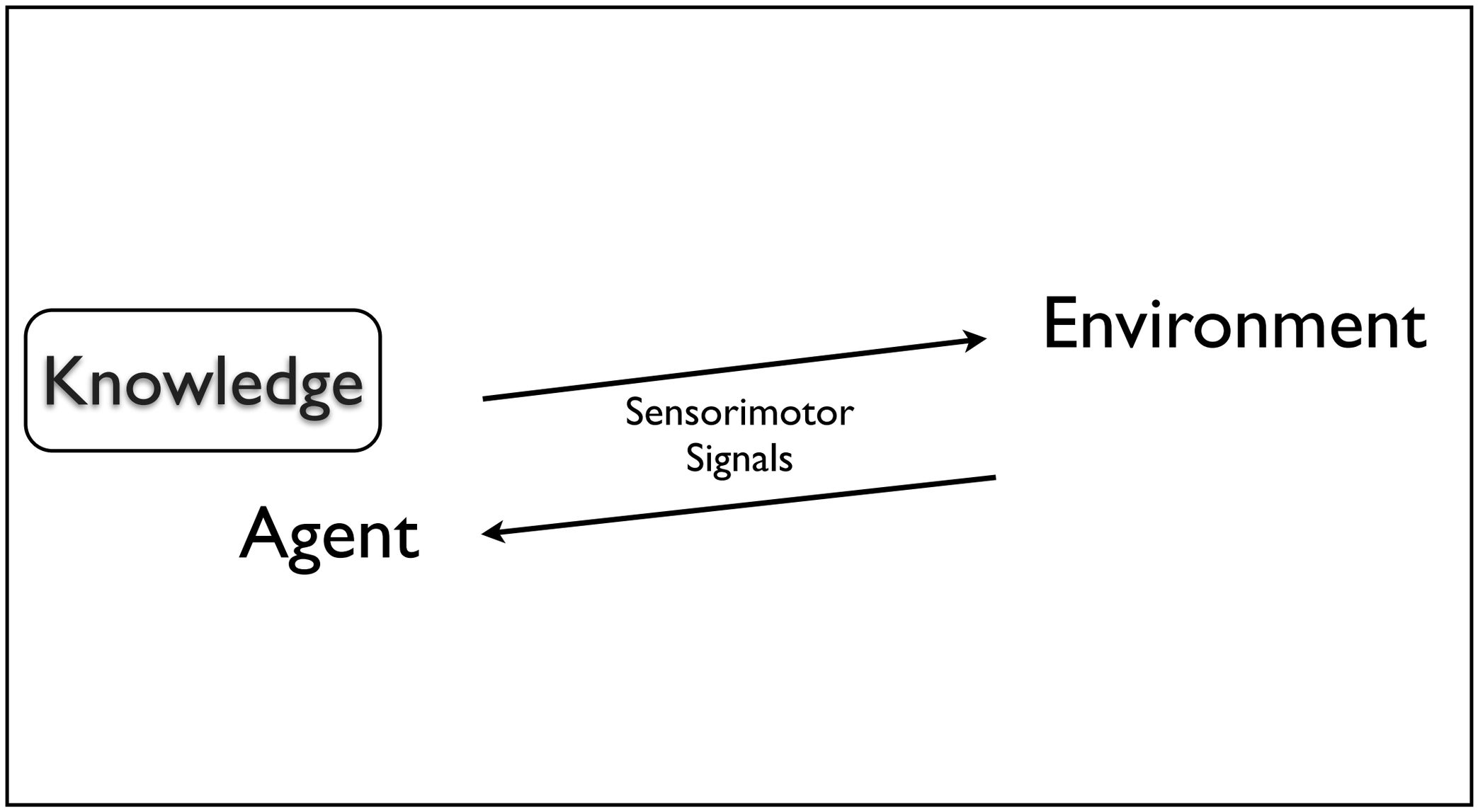
Verification



- Check for consistency
- Ask an oracle
- Check against data

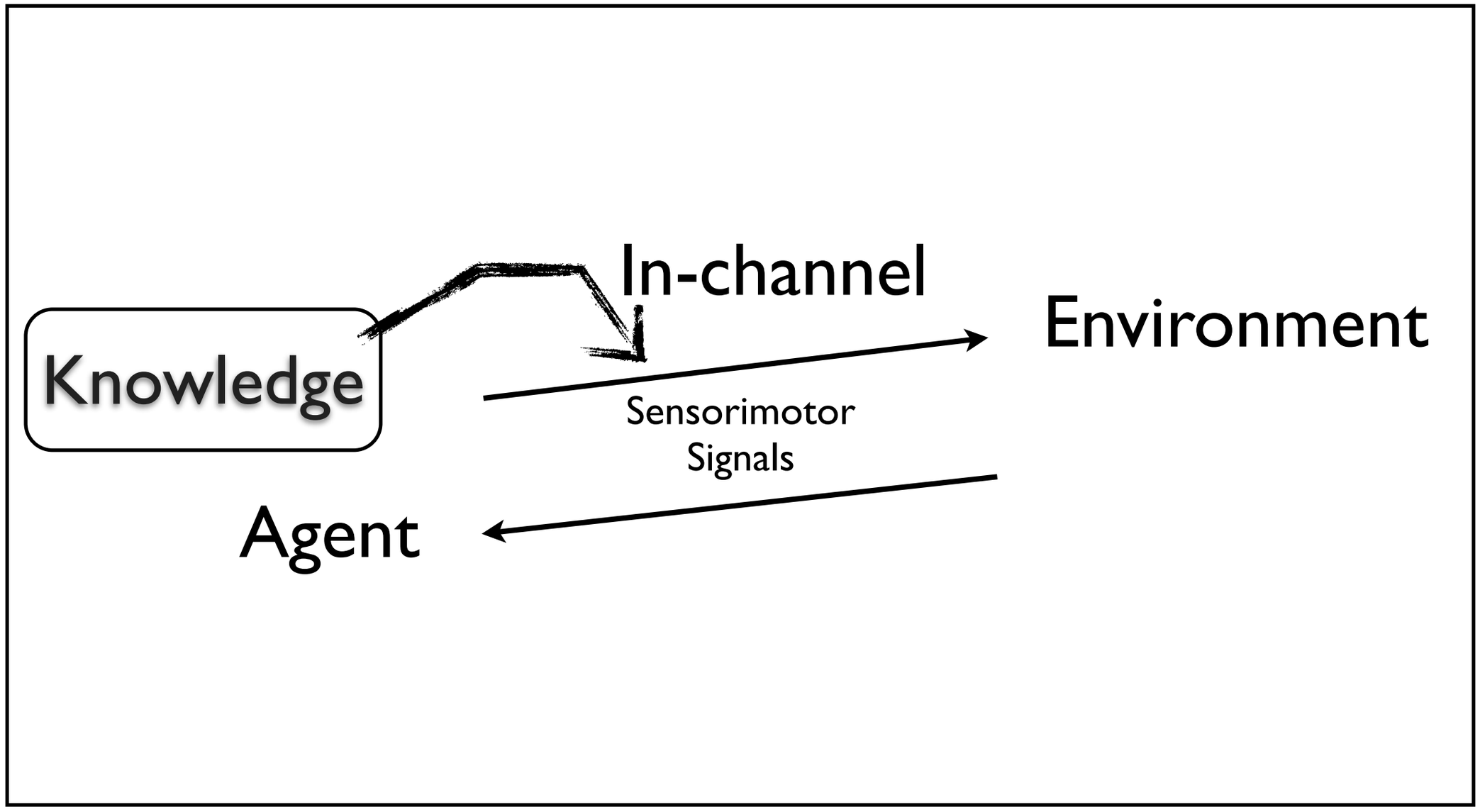
Compare with data---compare value to return
Hold out testing data

Approaches to Verification



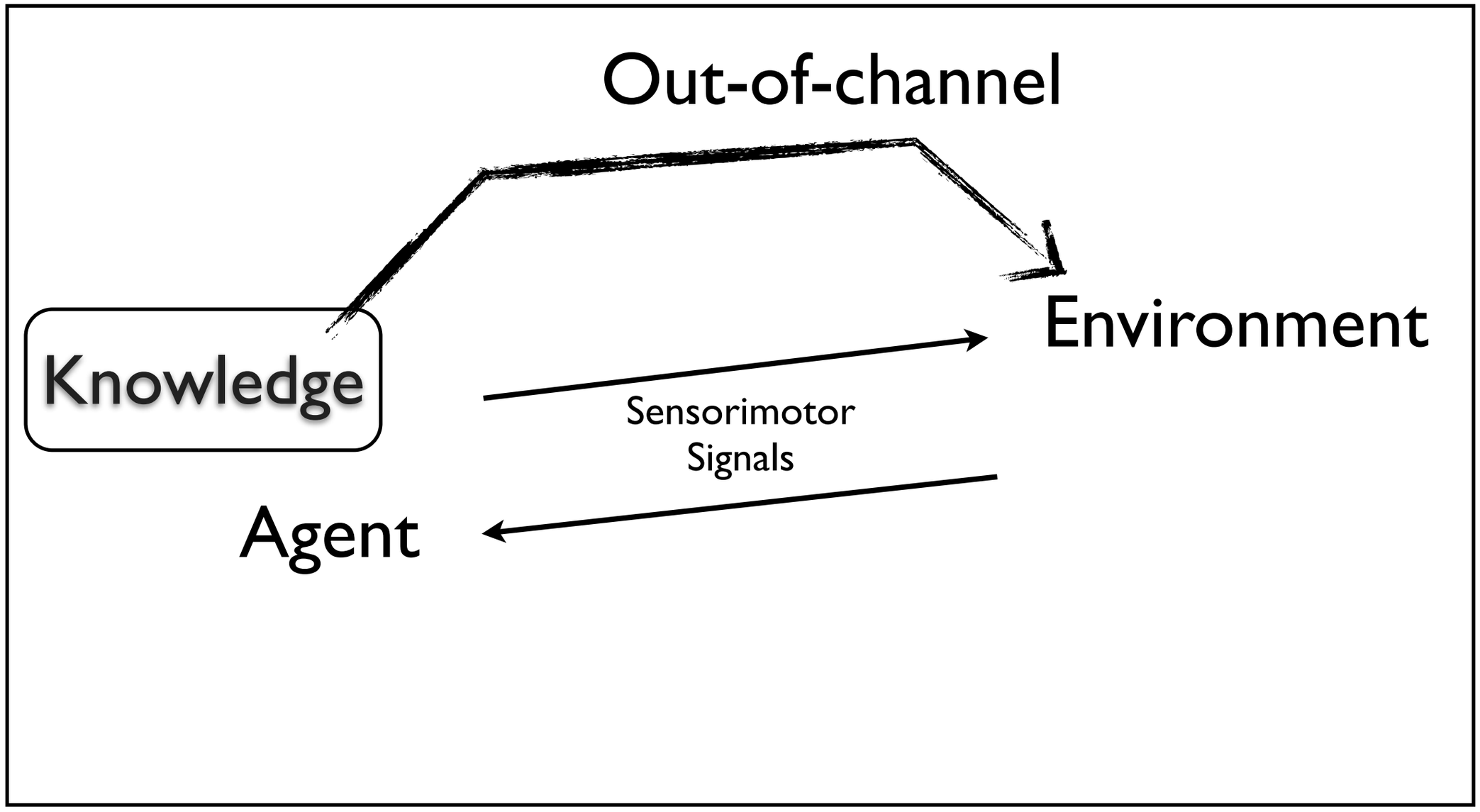
In-channel and out-of-channel

Approaches to Verification



In-channel and out-of-channel

Approaches to Verification

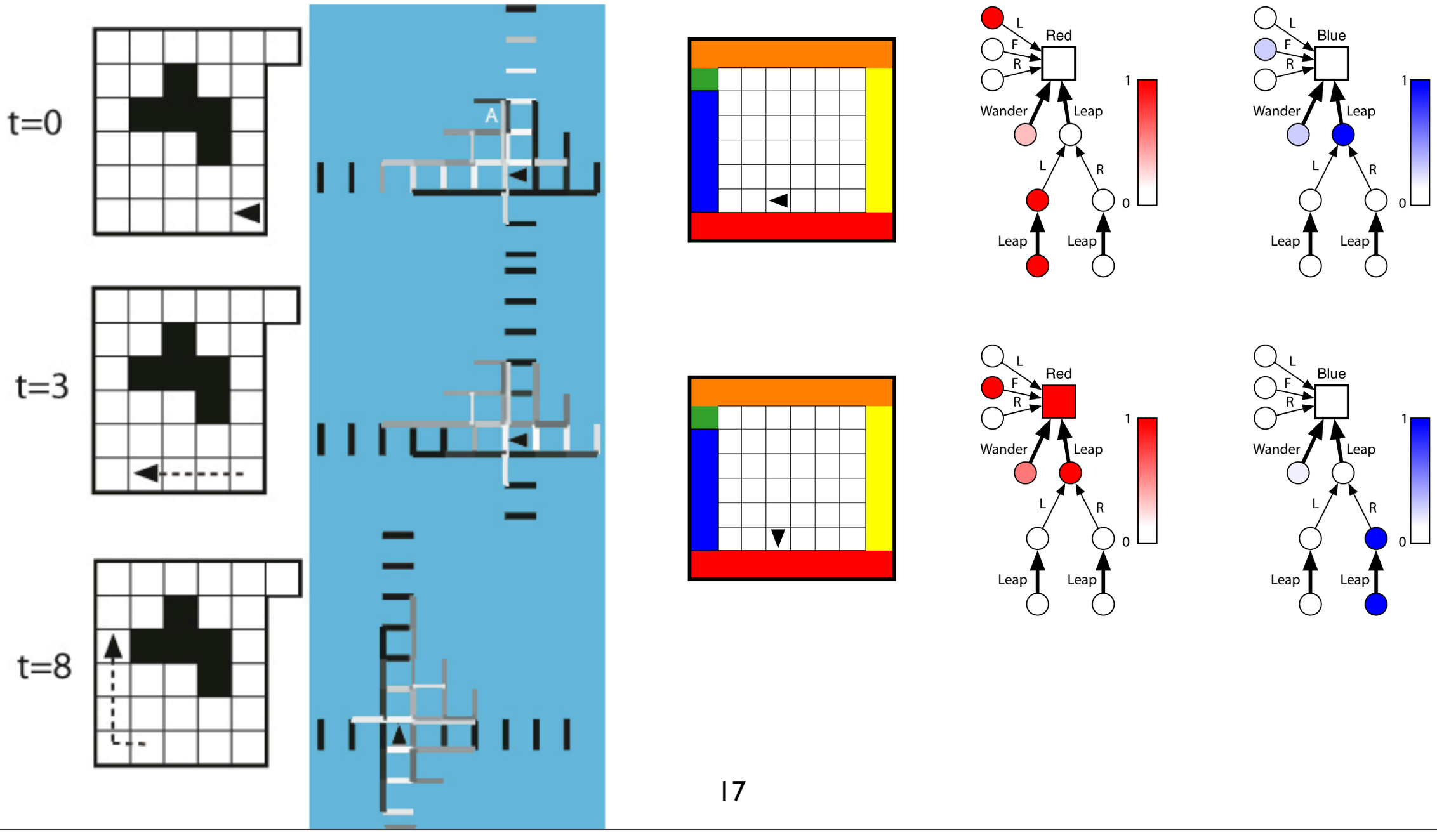


In-channel and out-of-channel

Reinforcement Learning

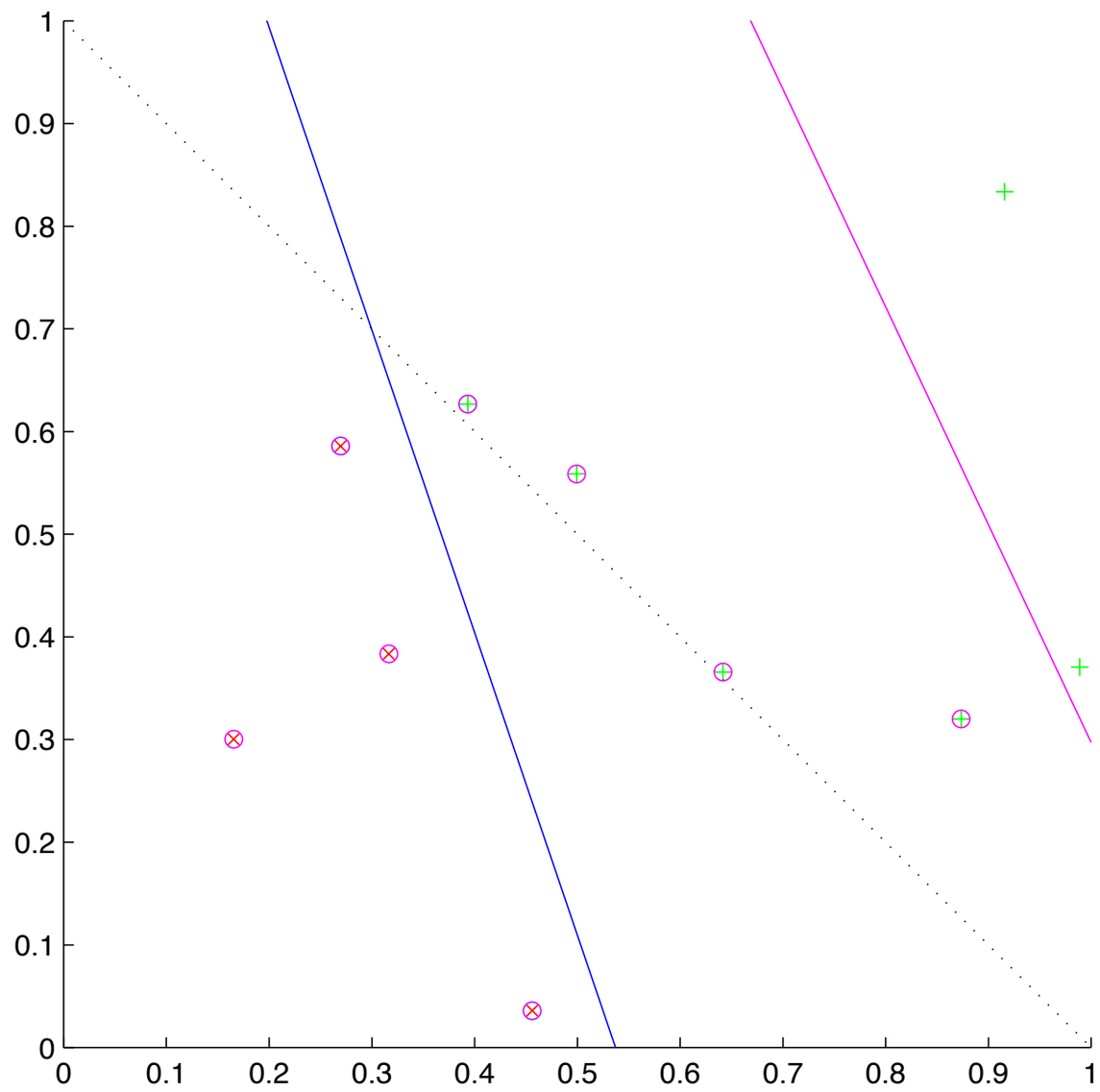
Can look at actual reward, even empirical expectation
But often find ourselves taking the out-of-channel perspective
Projected bellman error vs bellman error

Predictive Representations



Designed to be in-channel state representations
For validating our approaches we often look at out-of-channel information
Sometimes turn to oracles for learning too (compare state label to agent state)

Supervised Learning



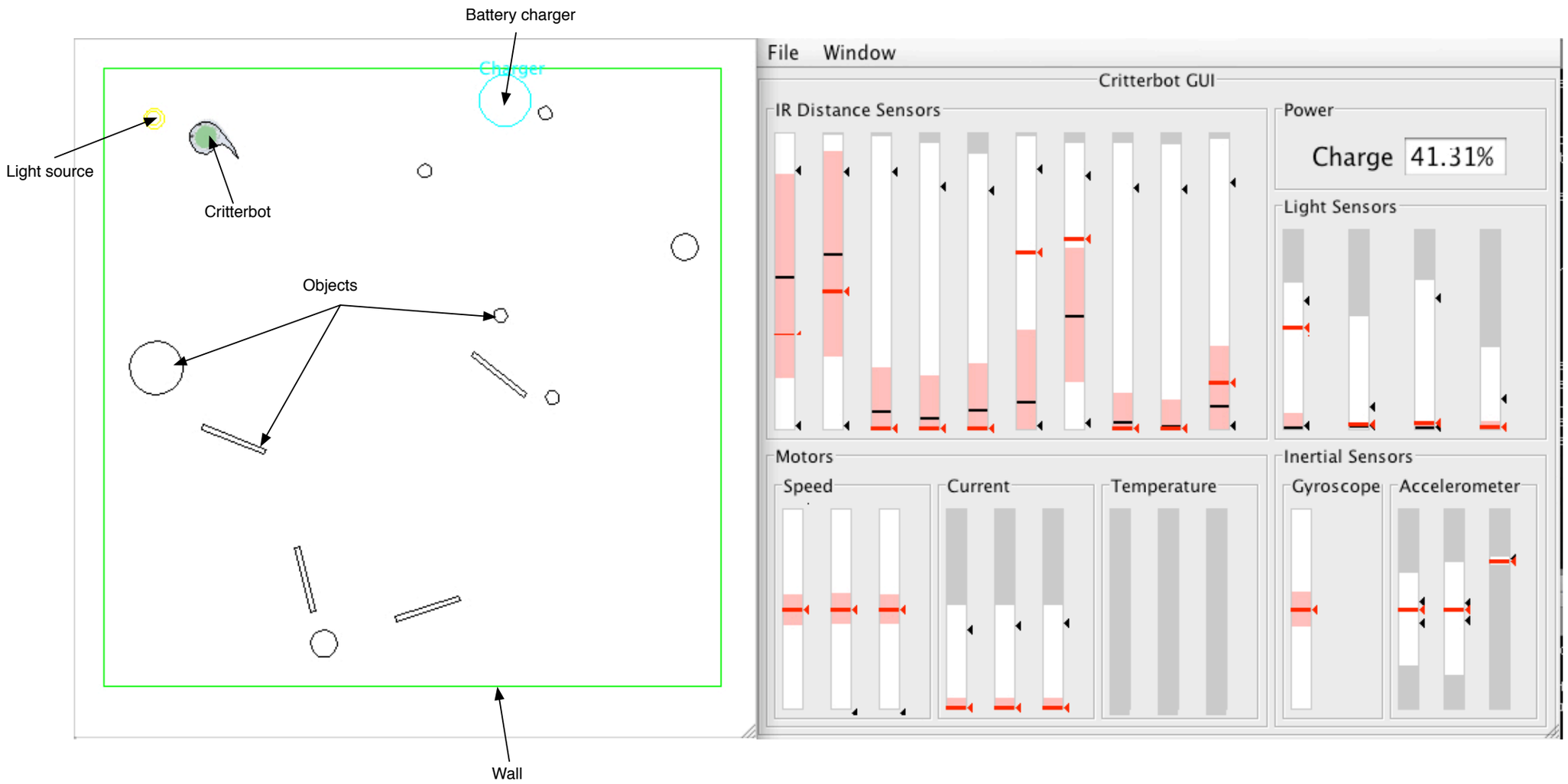
Can look at classification error on test data
Can compare the discriminate learned compared to that which generated the data.

Robotics

SLAM-trying to build a map, looking at inchannel information and using it to mold knowledge

Can ask an oracle, and we kinda need the validation tools

Robotics



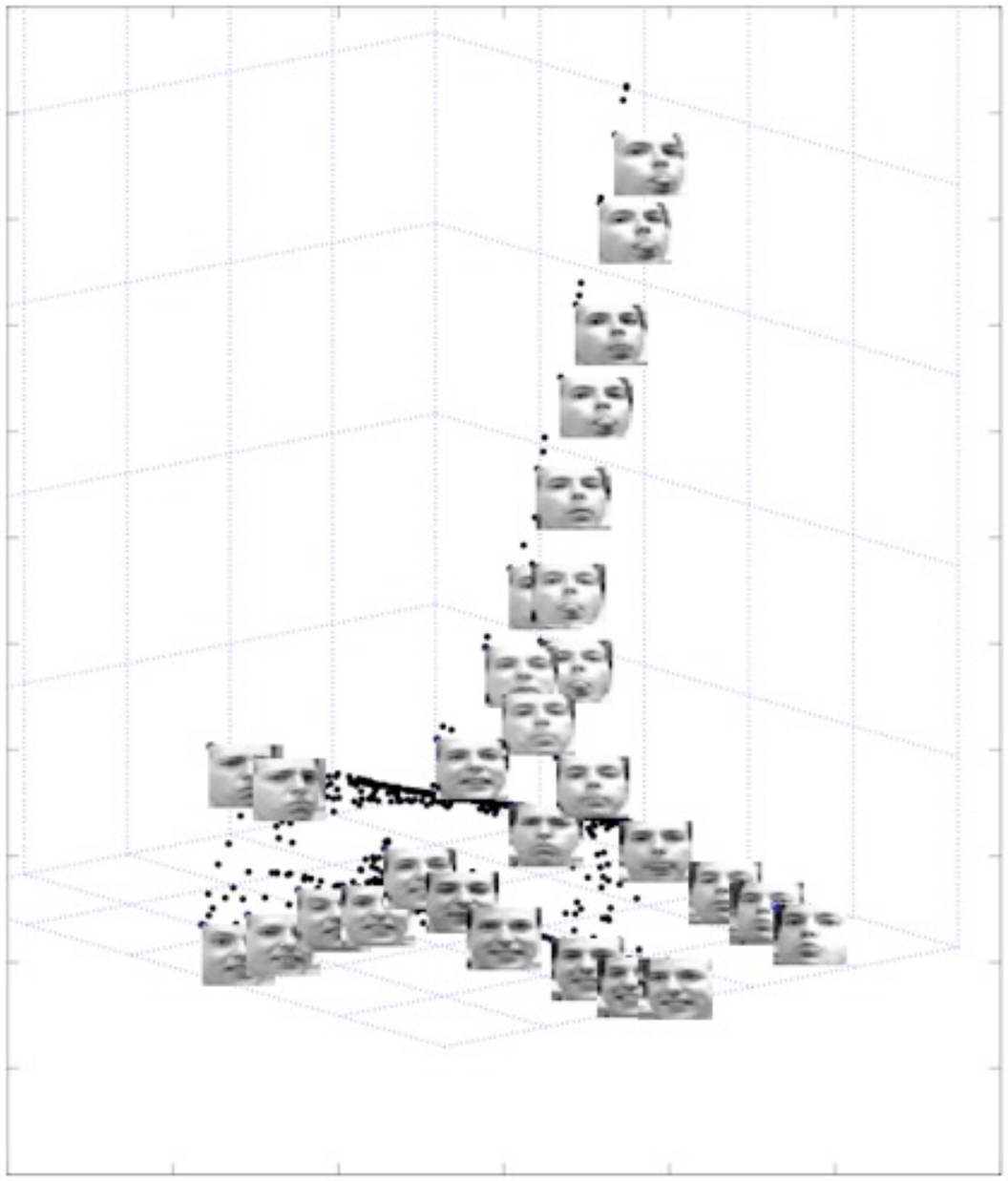
SLAM–trying to build a map, looking at inchannel information and using it to mold knowledge

Can ask an oracle, and we kinda need the validation tools

Manifold Discovery

Classic problem in manifold discovery, okay, you've minimized the loss function.
Did you get the right thing?

Manifold Discovery



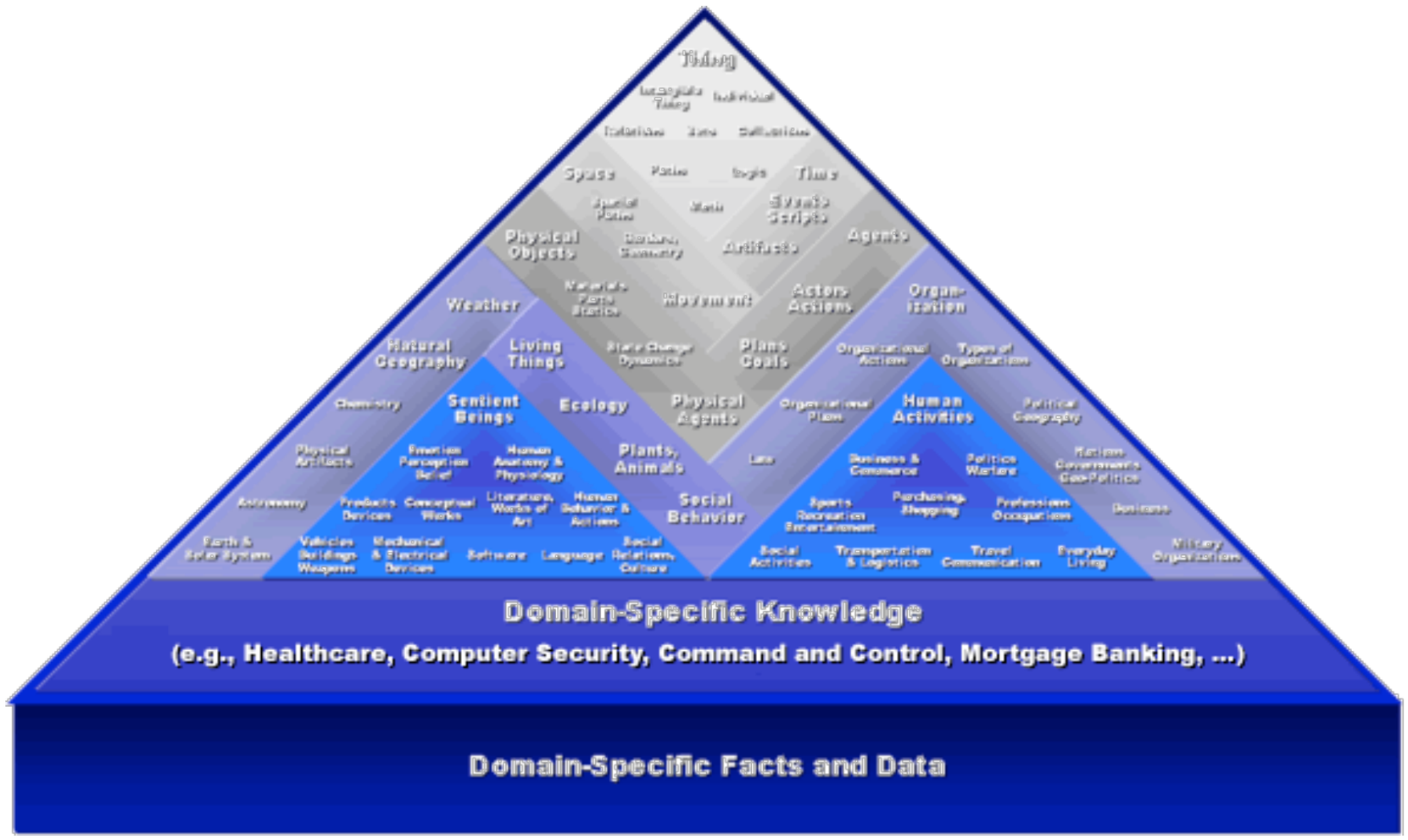
Classic problem in manifold discovery, okay, you've minimized the loss function.
Did you get the right thing?

Question-Answering Systems

21

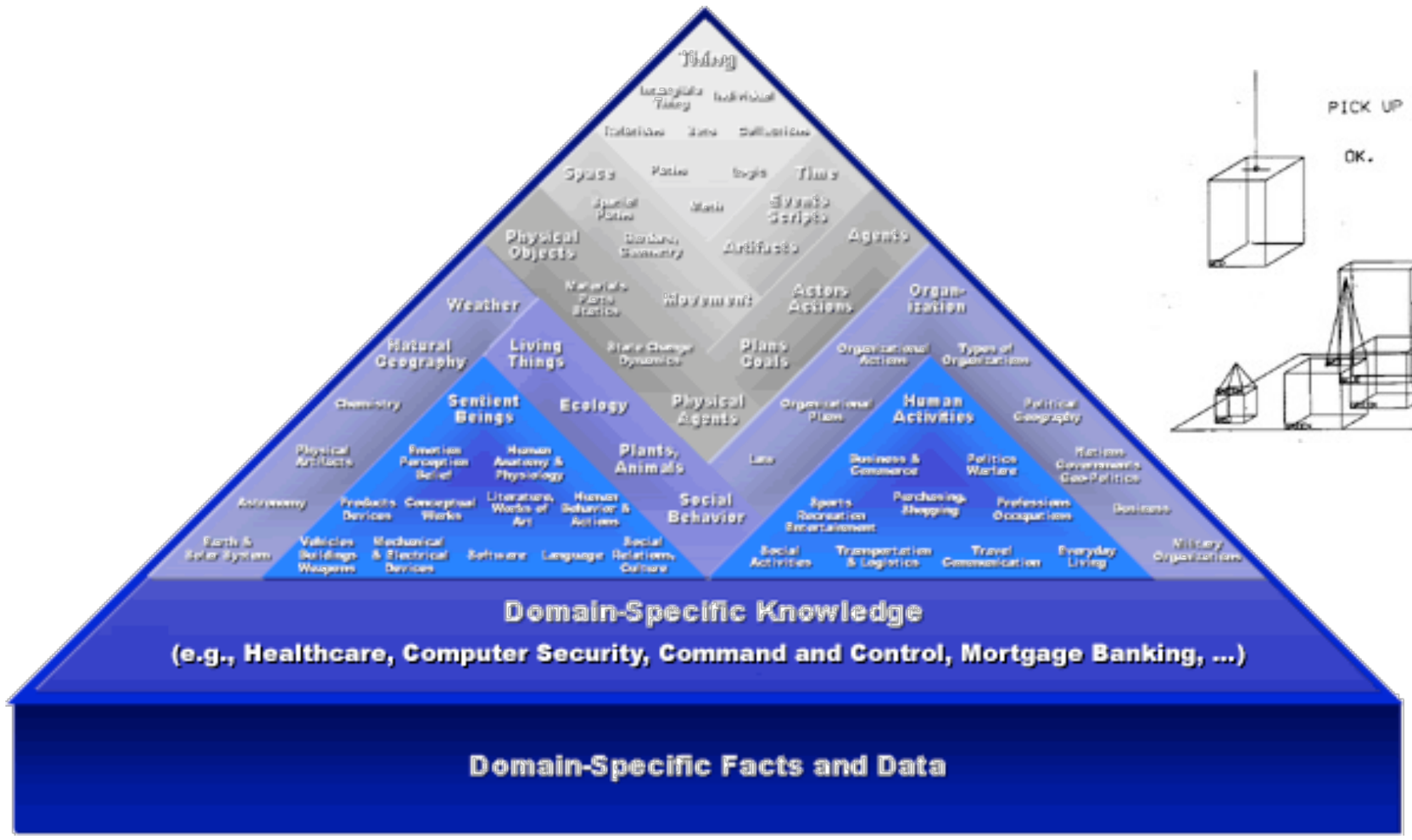
CYC---no in-channel verification possible
SHRDLU---does relate to data

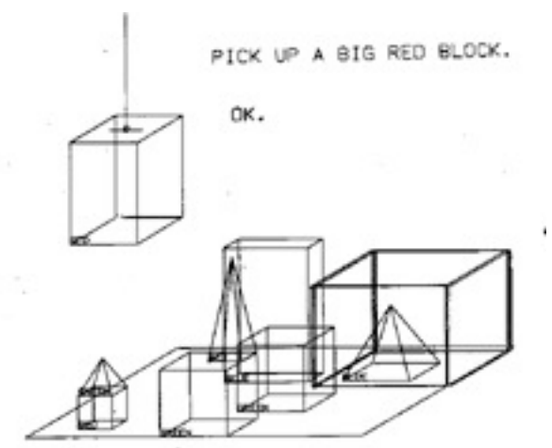
Question-Answering Systems



CYC---no in-channel verification possible
SHRDLU---does relate to data

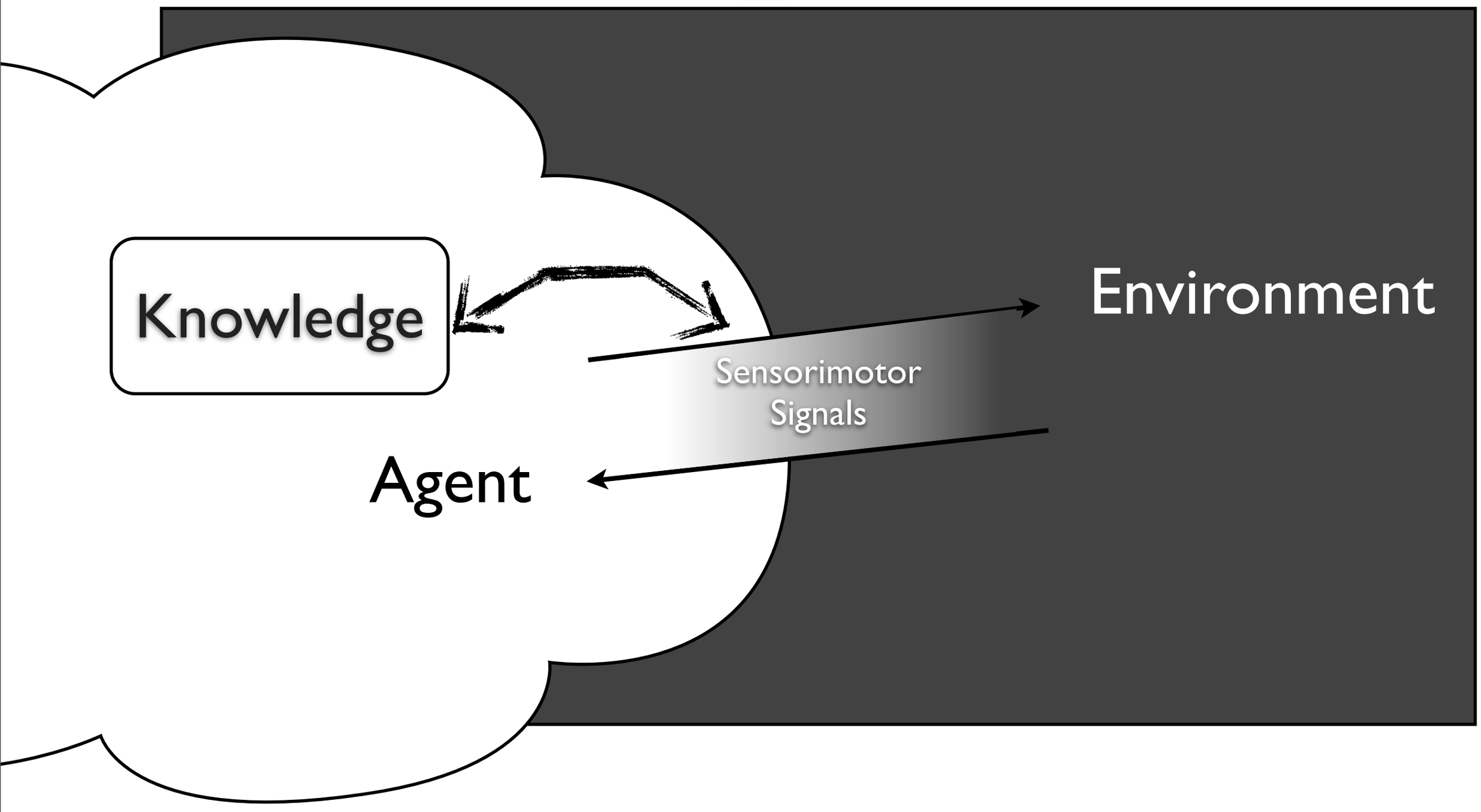
Question-Answering Systems





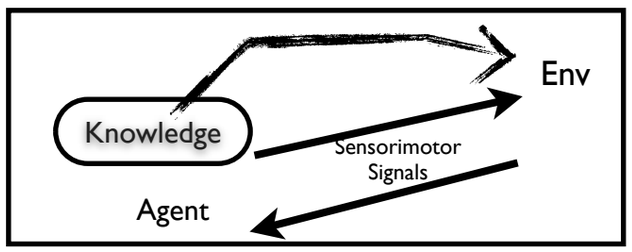
CYC---no in-channel verification possible
SHRDLU---does relate to data

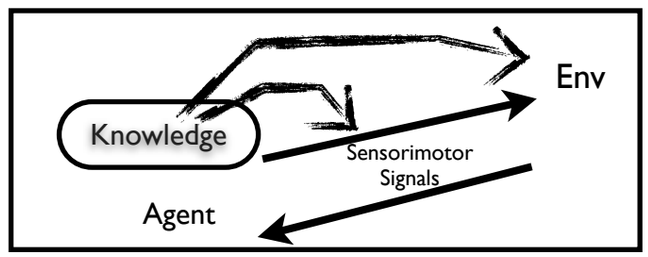
The Problem of Knowledge and Data

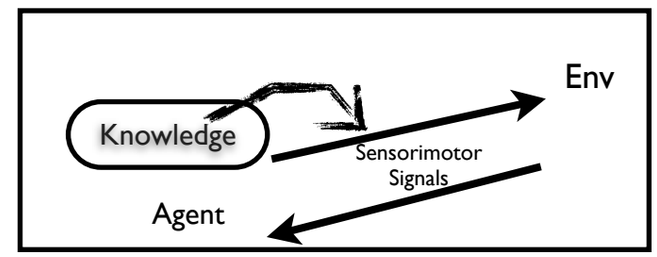


And this brings us to the problem of knowledge and data. How are they related?
- remember, problem in practice. Hard to make systems.
Matters for verification possibilities

Knowledge and Data



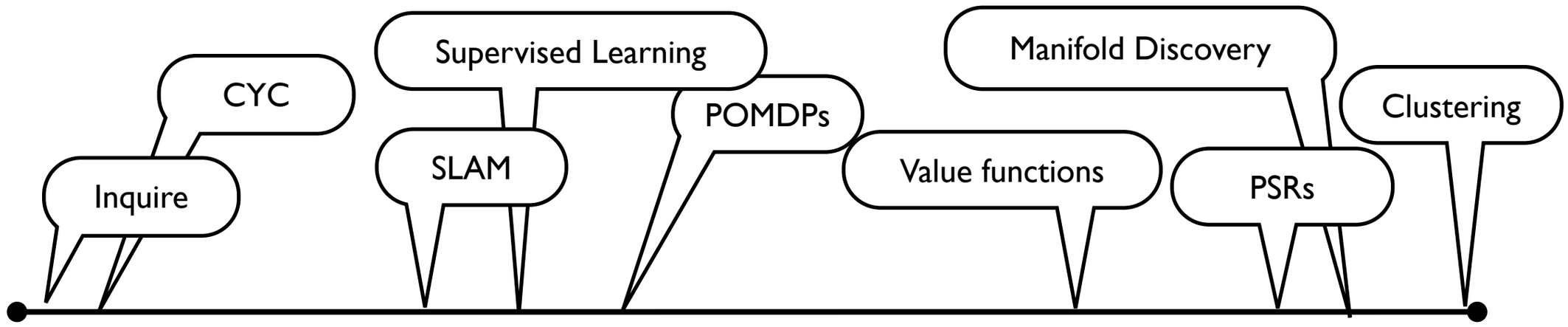




Objective Embedded Empirical

From a purely objective view to a purely empirical view
Definitely “it depends”

Approaches to Verification

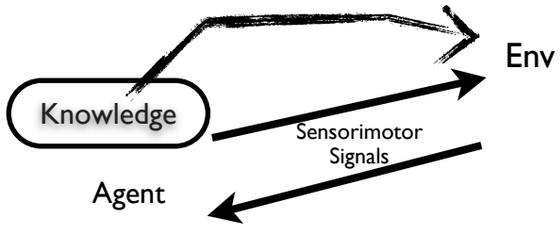


Objective Embedded Empirical

24

Can talk about where things fall on the spectrum

Objective



- Knowledge is about entities in the environment.
- Knowledge can be dissociated from data.
- Knowledge is verified out-of-channel

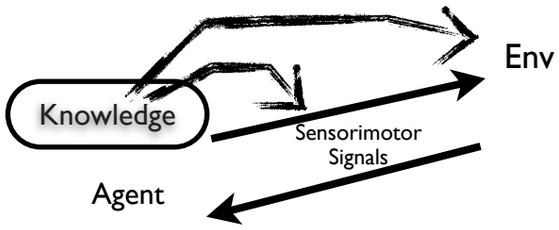
Objective

Knowledge is about entities in the environment.

Knowledge can be dissociated from data, reasoned about independently---no direct connection to data. Env entities generate data.

Knowledge can be checked for consistency and verified by oracle.

Embedded Systems

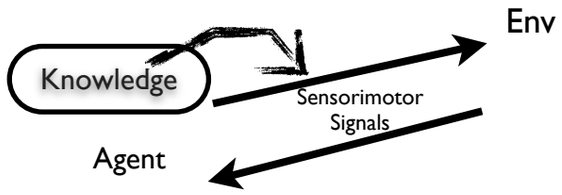


- Knowledge is about entities in the environment.
- Knowledge is updated according to data.
- Knowledge is verified in-channel and out-of-channel

Embedded

Knowledge is about entities in the environment.
Knowledge is updated according to data, k&d integrated
Knowledge can be checked against data and oracle

Empirical

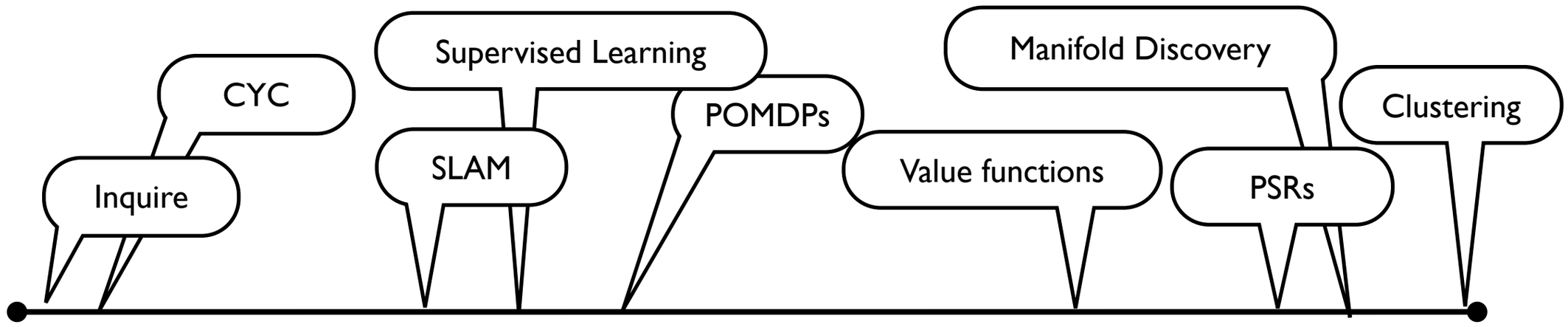


- Knowledge is about data.
- Knowledge can be dissociated from entities in the environment.
- Knowledge is verified in-channel.

Empirical
Knowledge is about data.
Connection intrinsic, k entirely data
Knowledge can be checked against data---no direct connection to entities in the env.
System generates data.

Why does this matter again? For in-channel verification. If the thing that gives knowledge its meaning and determines its veracity is in-channel

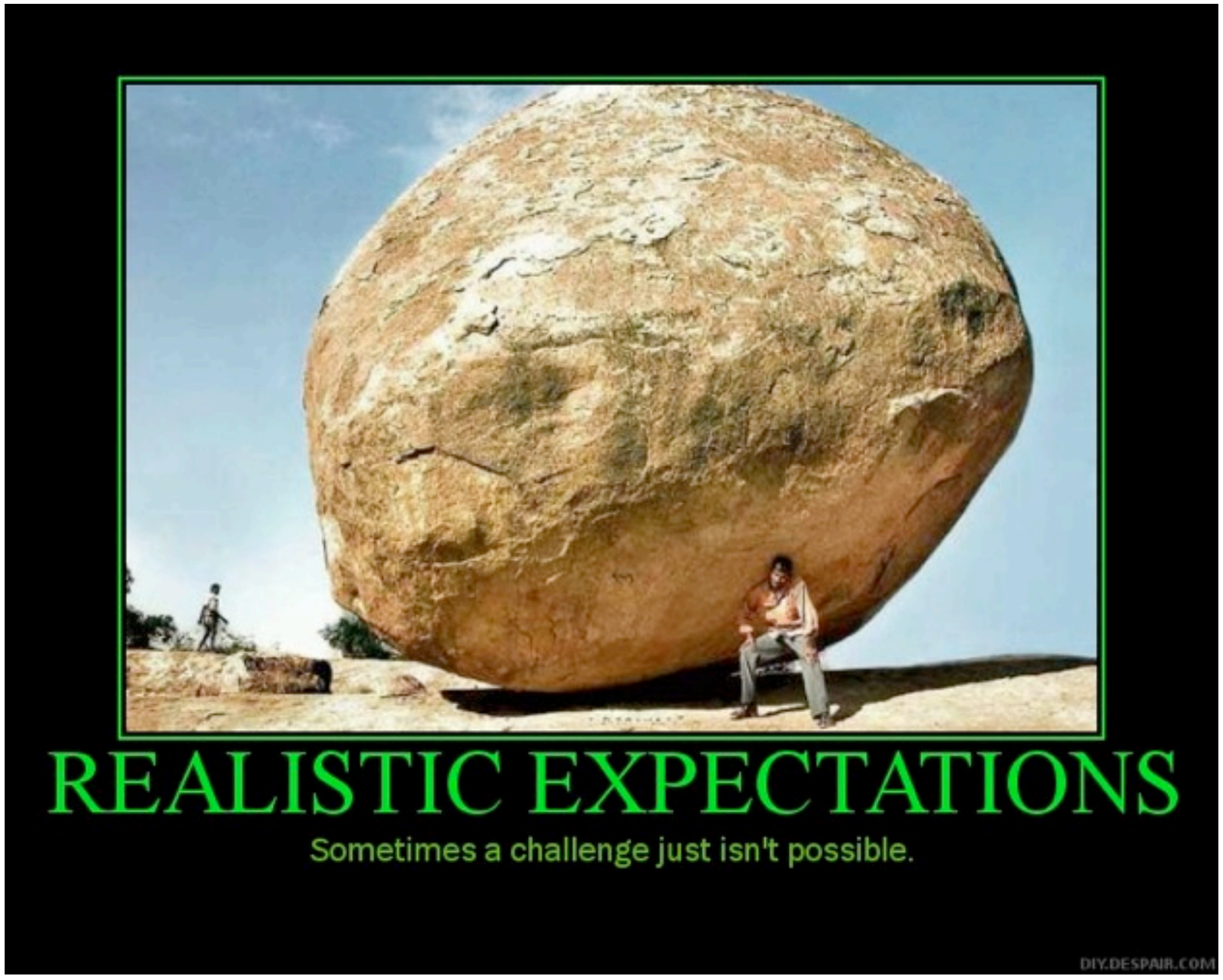
Approaches to Verification



Objective Embedded Empirical

Note: POMDPs and PSRs mathematically equivalent
Supervised learning and value functions can be either

The Research Project



29

What are the strengths and weaknesses of the extreme ends of the spectrum as discovered already?
What about this completely empirical approach?

Objective vs Empirical

- What have we seen as strengths and weaknesses?
  - Brittleness, scalability
  - Interpretability, usability

Objective---makes validation slightly easier, because you can see if the system knows what you want it to know, because it's constructed in human interpretable terms. On the other hand, in-channel verification is impossible because at its root the knowledge is not tied to accessible data, the oracle is required.
Empirical---makes validation difficult, problem of manifold discovery and data mining---got this structure, is it what we want? But verification can be entirely in-channel.

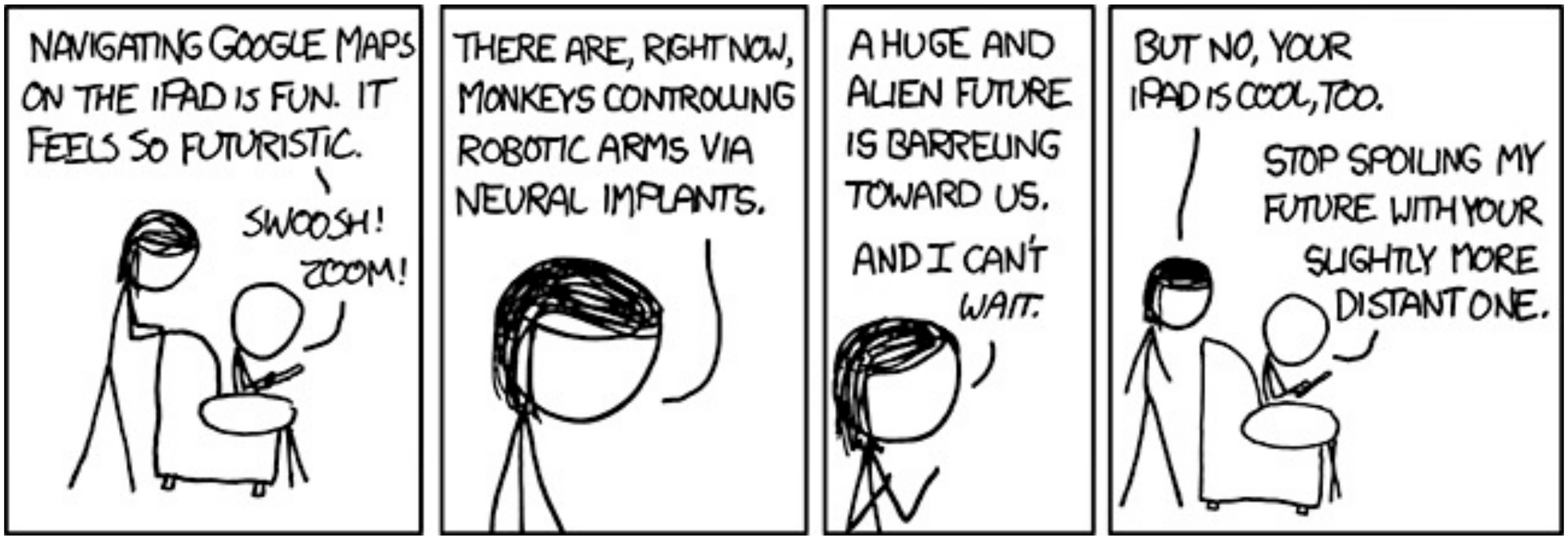
How does IT know it knows?

- Autonomous verification is great
- How do we say that this knowledge is appropriate?

31

"The mind does have to have a way of saying "hey, how am I doing?"
But then we want to be able to say this is useful, to validate it
Looking at how comp cog evaluates knowledge in non-linguistic species for ideas

Questions?


NAVIGATING GOOGLE MAPS ON THE IPAD IS FUN. IT FEELS SO FUTURISTIC.
SWOOSH!
ZOOM!
THERE ARE, RIGHT NOW, MONKEYS CONTROLLING ROBOTIC ARMS VIA NEURAL IMPLANTS.
A HUGE AND ALIEN FUTURE IS BARRELING TOWARD US.
AND I CAN'T WAIT.
BUT NO, YOUR IPAD IS COOL, TOO.
STOP SPOILING MY FUTURE WITH YOUR SLIGHTLY MORE DISTANT ONE.