Under the Hood

The "Generality" of an AI

How many tasks can the AI perform?
What kinds of tasks can an AI perform

From Russell & Norvig (2010) "*Artificial Intelligence: A Modern Approach*",

Task Environment	Observable	Agents	Deterministic	Episodic	Static	Discrete
Crossword puzzle	Fully	Single	Deterministic	Sequential	Static	Discrete
Chess with a clock	Fully	Multi	Deterministic	Sequential	Semi	Discrete
Poker	Partially	Multi	Stochastic	Sequential	Static	Discrete
Backgammon	Fully	Multi	Stochastic	Sequential	Static	Discrete
Taxi driving	Partially	Multi	Stochastic	Sequential	Dynamic	Continuous
Medical diagnosis	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Image analysis	Fully	Single	Deterministic	Episodic	Semi	Continuous
Part-picking robot	Partially	Single	Stochastic	Episodic	Dynamic	Continuous
Refinery controller	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Interactive English tutor	Partially	Multi	Stochastic	Sequential	Dynamic	Discrete

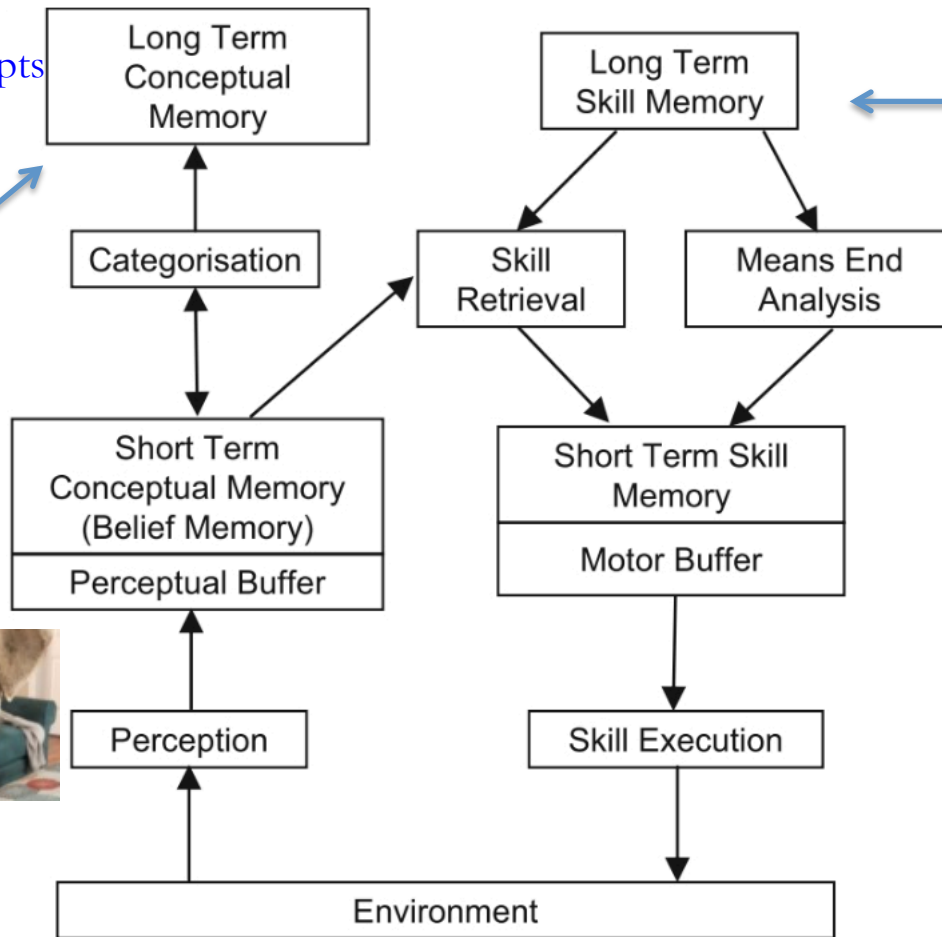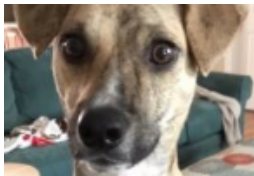**Figure 2.6** Examples of task environments and their characteristics.

<Fully, Single, Deterministic, Episodic(Bernouli), Static, Discrete> the "simplest"
<Partial, Multi, Stochastic, Sequential(Unbounded), Dynamic, Continous> the "hardest"

Cognitive architectures to implement general AI

Doug's password
Words, definitions, concepts
Mr Cuddles is a terrier
  kind of dog
  kind of mammal
  kind of vertebrate
  kind of animal
(semantic web)

Calvin voice samples
Calvin voice invariants
…

Typing Doug's password
Constructing sentences
Turning a wheel (degree left or right)
Applying a brake

Buy ticket (from X to Y)
Drive route (from X to Y)

Parameters, X and Y

Sequential memory?

Long Term Conceptual Memory → Categorisation ↔ Short Term Conceptual Memory (Belief Memory) / Perceptual Buffer ← Perception ← Environment

Long Term Skill Memory → Skill Retrieval / Means End Analysis → Short Term Skill Memory / Motor Buffer → Skill Execution → Environment

Icarus architecture (Langley et al;
Image from Chong et al ,
adapted from Langley, 2004)

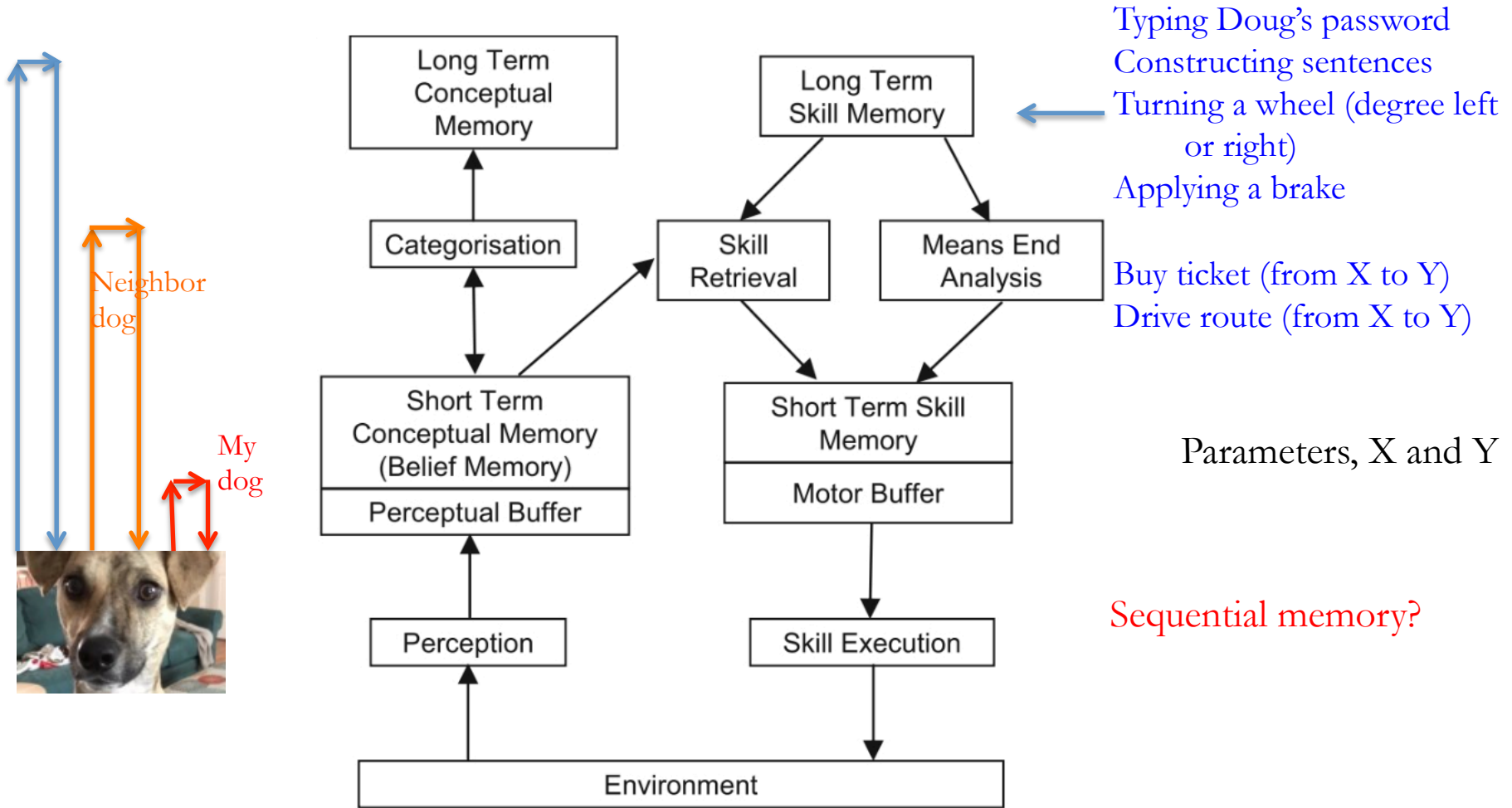# Cognitive architectures to implement general AI

Typing Doug's password
Constructing sentences
Turning a wheel (degree left or right)
Applying a brake

Buy ticket (from X to Y)
Drive route (from X to Y)

Parameters, X and Y

Sequential memory?

Neighbor dog

My dog

Icarus architecture (Langley et al;
Image from Chong et al ,
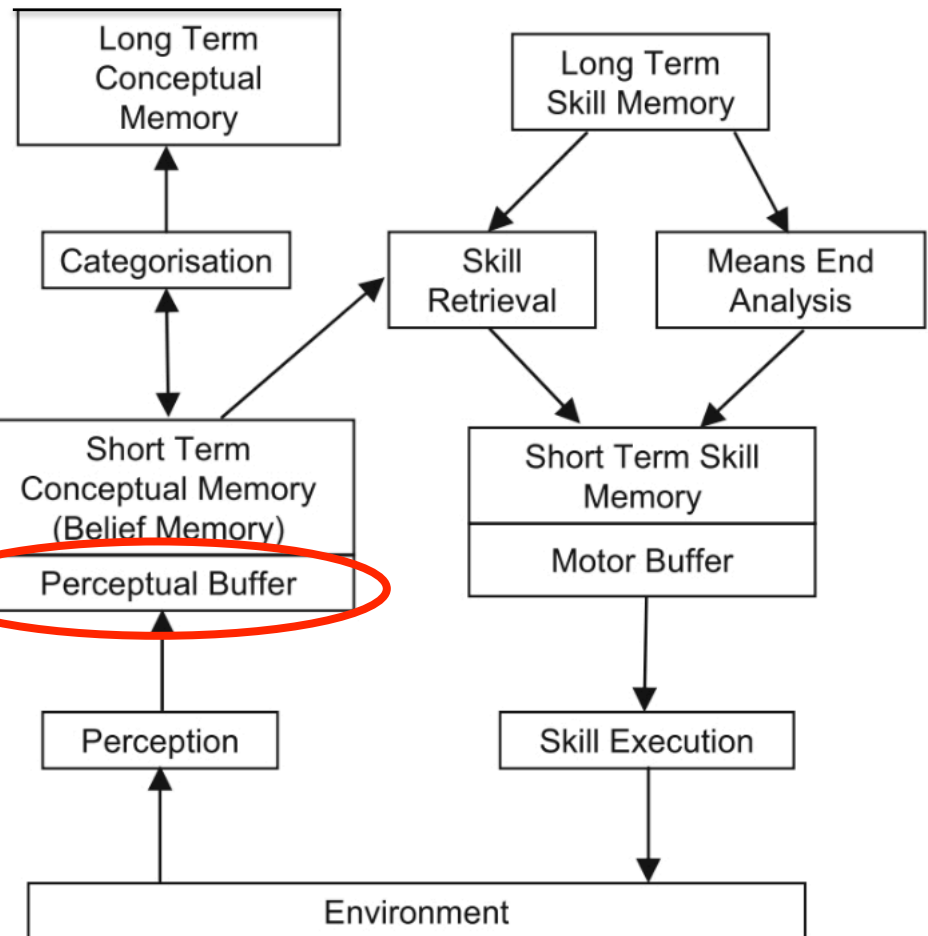adapted from Langley, 2004)

Calvin nodded. She said, quietly, "Elvex, you will not move nor speak nor hear us until I say your name again."

If Elvex doesn't "hear us" then how can it hear its name?

IF "Elvex" (by Calvin)
   THEN awaken (other) Elvex modules
   ELSE delete percept

Calvin voice
Retrieved and put
in buffer with label

Icarus architecture (Langley et al;
Image from [Chong et al](#) ,
adapted from Langley, 2004)

Long Term Conceptual Memory

Long Term Skill Memory

Categorisation

Skill Retrieval

Means End Analysis

Short Term Conceptual Memory (Belief Memory)

Perceptual Buffer

Short Term Skill Memory

Motor Buffer

Perception

Skill Execution

Environment

Messing with an AI

Calvin turned to Rash "I am going to wake Elvex up now. Do not say anything about the possibility of my turning its positronic brain into an ingot. Do you understand?" Rash nodded.

"Elvex, wake up. Do you hear me?"

Elvex then ….

Humanist high-level "agent program"
Represented as a flowchart with functional stubs

http://mchrbn.net/ethical-autonomous-vehicles/

Unexpected situation detected

Retrieve automated car's information

Detect all objects in the scene that could potentially be involved with the decision-making process

Store objects and their information

Use the initial output (no action taken by the car)

Run simulation of damages received by each human with current output

Search for alternatives. Is there another output?

YES

Transmit the chosen output to the car so it acts accordingly

Compare all outputs and select the one with the least overall damages to humans

Store damages for current output

Keep current dataset

YES

Do all outputs involve at least one death?

NO

Purge outputs that include death

NO