

Logic and Intelligence

Carolyn Talcott

LAP

September 2026

Logic vs Intelligence

- Formal logic was developed by ancient philosophers to analyze discourse and determine if an argument is valid or not.
- A formal logic provides a language for stating propositions and arguments, and rules for deriving consequences and checking validity of arguments/derivations.
- (Natural) intelligence is the ability of living creatures to survive in a variety of conditions, to achieve goals, to plan, reason, learn, adapt.
- Intelligence evolved biologically to meet changing environmental conditions (over millions of years!).

Logic and Intelligence

- Two different notions of reasoning.
 - Static, reasoning in the abstract, a closed system
 - Dynamic, reasoning in the world. sensing and acting, evolving
- Mechanized reasoning system (MRS)
 - a logic system embedded in a mechanism that supports interaction
 - sensing/input — ideas, physical environment
 - acting/output — evolving reasoning state, affecting the environment
- Can an MRS be endowed with properties of evolving intelligence?
 - what are the building blocks
 - what are meaningful levels of competence/intelligence

Plan

Towards understanding building blocks for building mechanized intelligence:

I. Summary of key mechanisms and features of intelligence

- based on 'A Brief History of Intelligence' by Max Bennet

II Description of the FOL system

- an early mechanized reasoner with some mechanisms for 'intelligence'

III Summary / Questions / Challenges

How does intelligence / a mind work?

Questions motivating the FOL project

- How can a mind forget?
- How can a mind be interrupted?
- How does perceptual information get to be cognitive information?
- How we can comprehend infinite objects such as the set of numbers?
- How we can reason about others knowledge / motivations?
- What does it mean to “know” in an ever changing environment?

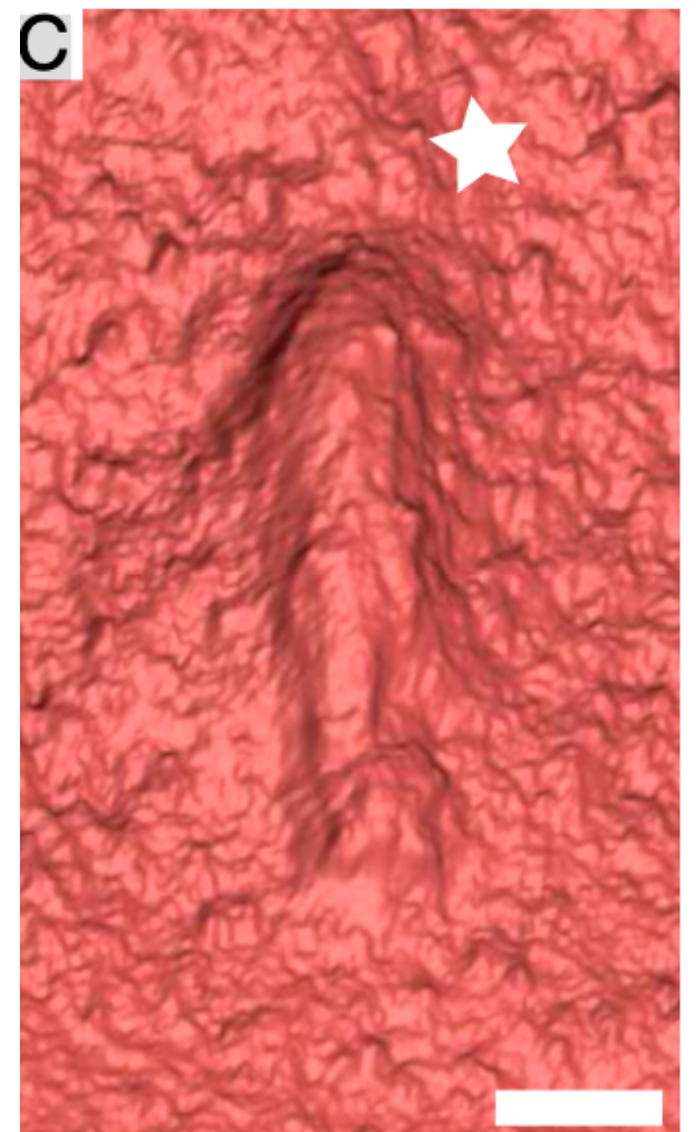
Stages of Intelligence

Key Stages of Brain Development: Bennet's 5 Breakthroughs

1. [-500m years/bilaterals] steering / simple sensor/actuator network with small central processor
2. [-300-400m years/vertabrates] basic reinforcement learning / brain developing structure
3. [-100m years/mamals] simulation of world/model based learning / neocortex emerges
4. [-30m years/primates] ability to model mind (self and others), new regions of neocortex
5. [-100k years/human] language/labeling enabling exchange and persistence of knowledge / bigger brain, new regions, more complex brain wiring

Breakthrough 1 (bilateral) steering

- Navigating by categorizing stimuli as good or bad—turn toward good, turn away from bad.
- Features/mechanisms
 - Basic neuron net : observe, evaluate, act
 - sensor neuron -> valence neuron -> motion neuron
 - valence neurons categorize signal,
 - central neurons combine valence values
 - Add sensing chemicals -- provide context
 - spike signal is short
 - chemical signals last and spread
 - decisions becomes context dependent
 - if hot then cool is good, if cold then hot is good



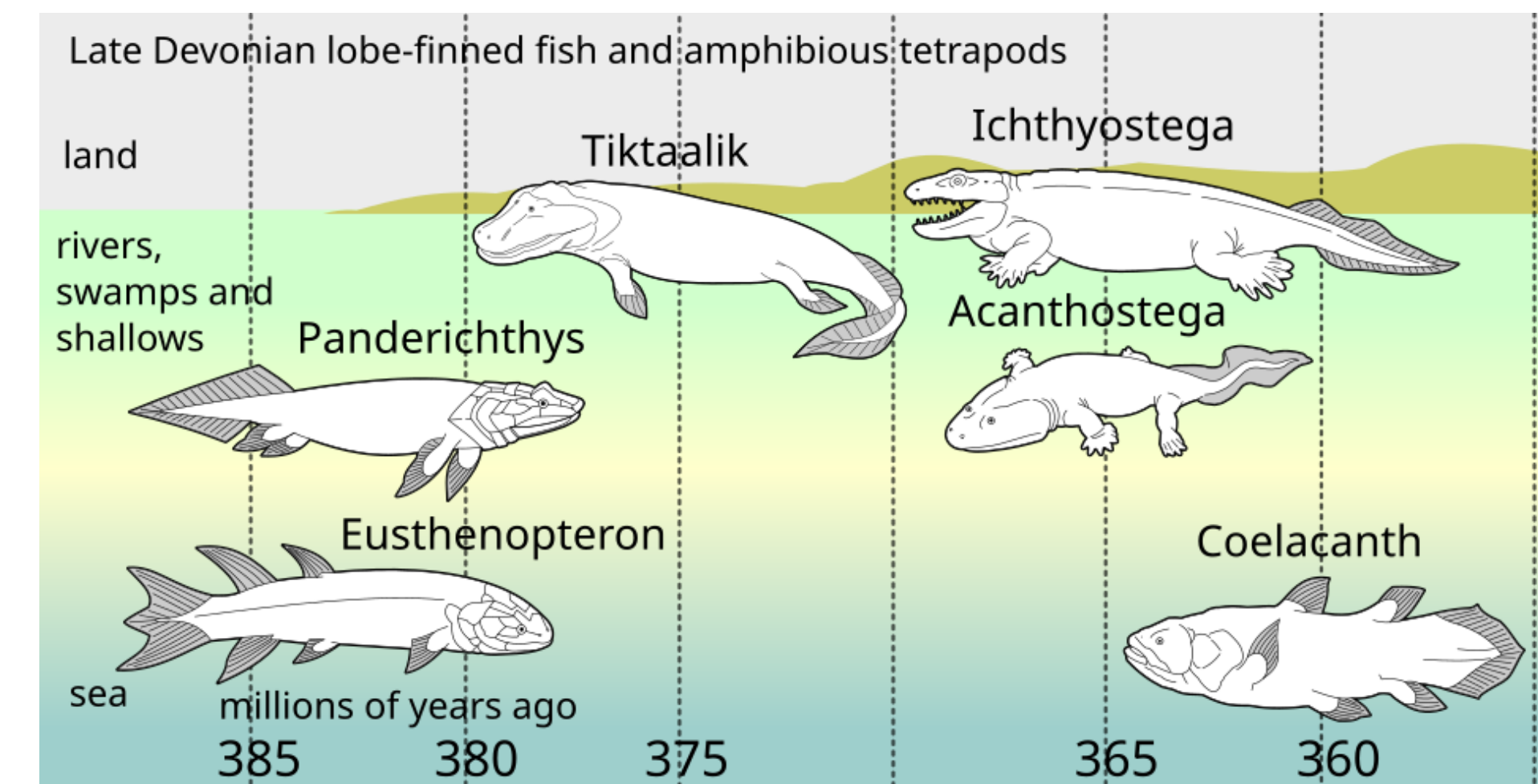
Breakthrough 1 — continued

- Chemicals also encode affect
 - Dopamine ~ want, motivates acquisition
 - Serotonin ~ lack of want, conserve energy when escape not possible
 - Adrenaline ~ stress, focus energy on response
 - Opioids ~ anti stress, enhance recovery
- Associative learning -- basic reflex reaction
 - rewiring of neuron network, increase/decrease connection strength
 - change notion of good/bad in response to changing world

Breakthrough 2 (vertebrates) Reinforcement learning

The brain evolves more structure

- cortex -- pattern recognizer, support for generalization
 - lateral -- recognize smells
 - ventral (amygdala) -- visual/audio patterns
 - medial (hippocampus) -- spatial maps, self location on map, distinguish self and world
- basal ganglia (BG) -- trial and error learning
 - timing cells -- timing of perception, when to act
 - learning from omission
- hypothalamus (from bilateral brain) houses valence machinery



Breakthrough 2 continued

- Reinforcement learning strategy -- actor-critic system
 - actor (BG) does experiments/actions
 - critic (Hypothalamus, valence neurons) rewards moves to better positions by changing dopamine level
- actor-critic system learns to repeat behaviors that lead to positive valence and inhibit behaviors that lead to negative valence
- can learn simple sequences of actions

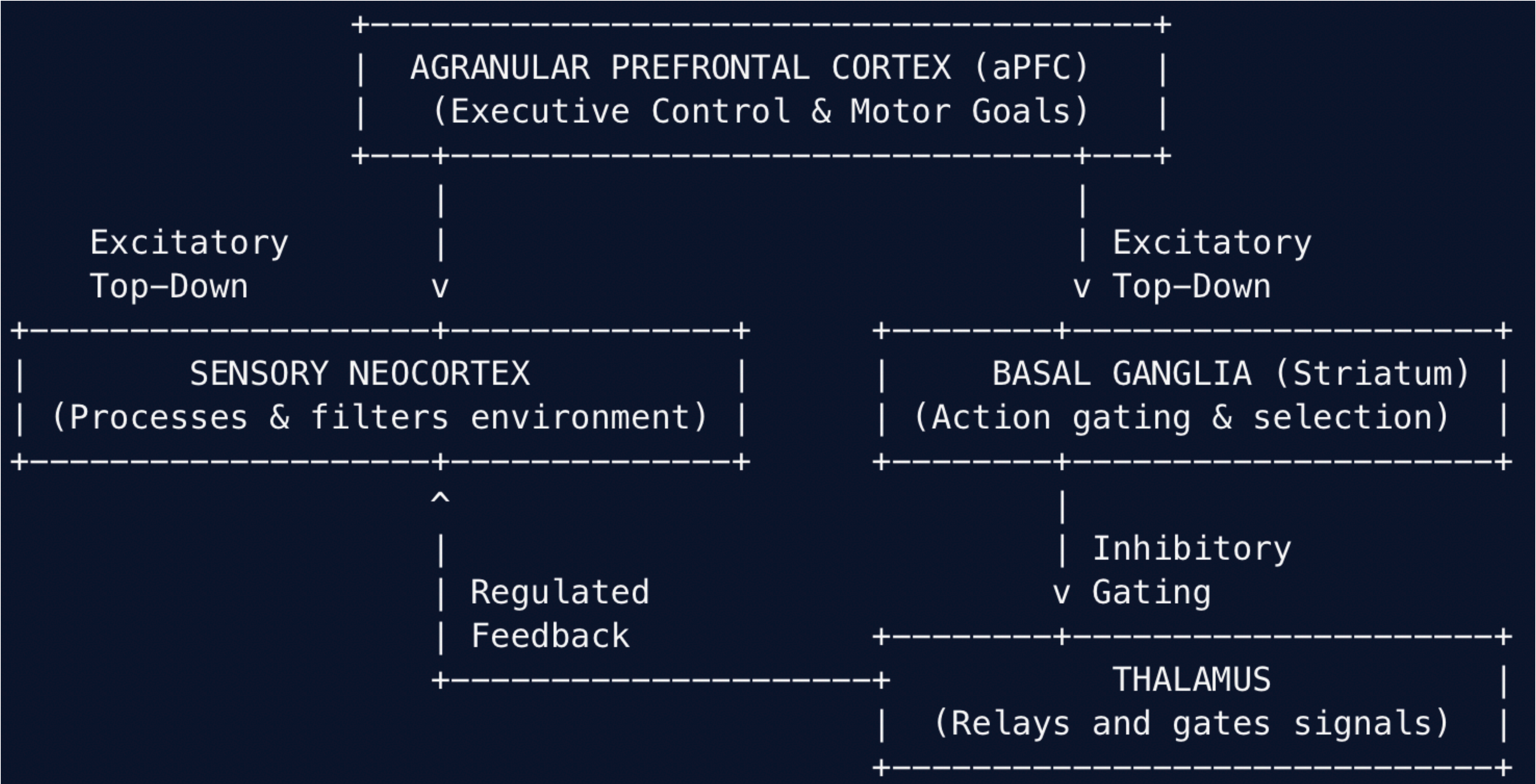
Breakthrough 3 (mammals) Simulation

Mentally simulating stimuli and actions — world model

- Enabled by appearance of the neocortex (NC)
 - sensory NC — model building/simulation—explains external world
 - agranular prefrontal NC (aPFC) —decides when and what sNC should simulate
- NC consists of many identical 6 layer columns of neurons
 - NC function/behavior depends on input and output wiring
 - Is the NC turing complete?
- Motor cortex provides fine motor control, complex movements



Breakthrough 3 — control of simulation



Breakthrough 3 new abilities

Abilities enabled by the neocortex simulation ability

- High level goals modeled in aPFC (deliberates) vs low level goals modeled BG (reactive) — thinking fast (BG) vs thinking slow (aPFC)
- Model of external world (sensory input) vs model of self (aPFC input)
- Trial and error experiments for making decisions
- Counterfactual learning -- imagining alternate pasts, distinguishing between correlation and causation
- Episodic Memory (memory of specific past events) — simulated recreation of past events, may get it wrong!

Breakthrough 4 (primates) modeling minds

- The granular prefrontal cortex region of NC (gPFC) appears
- The gPFC models/explains self, intent, predicts self future, model other minds (theory of mind).
- gPFC learns more abstract generative models (2nd order) of aPFC driven behavior -- explains intent
- How it works
 - aPFC gets input from the amygdala and hippocampus -- modeling the world
 - gPFC gets input from the aPFC -- modeling self

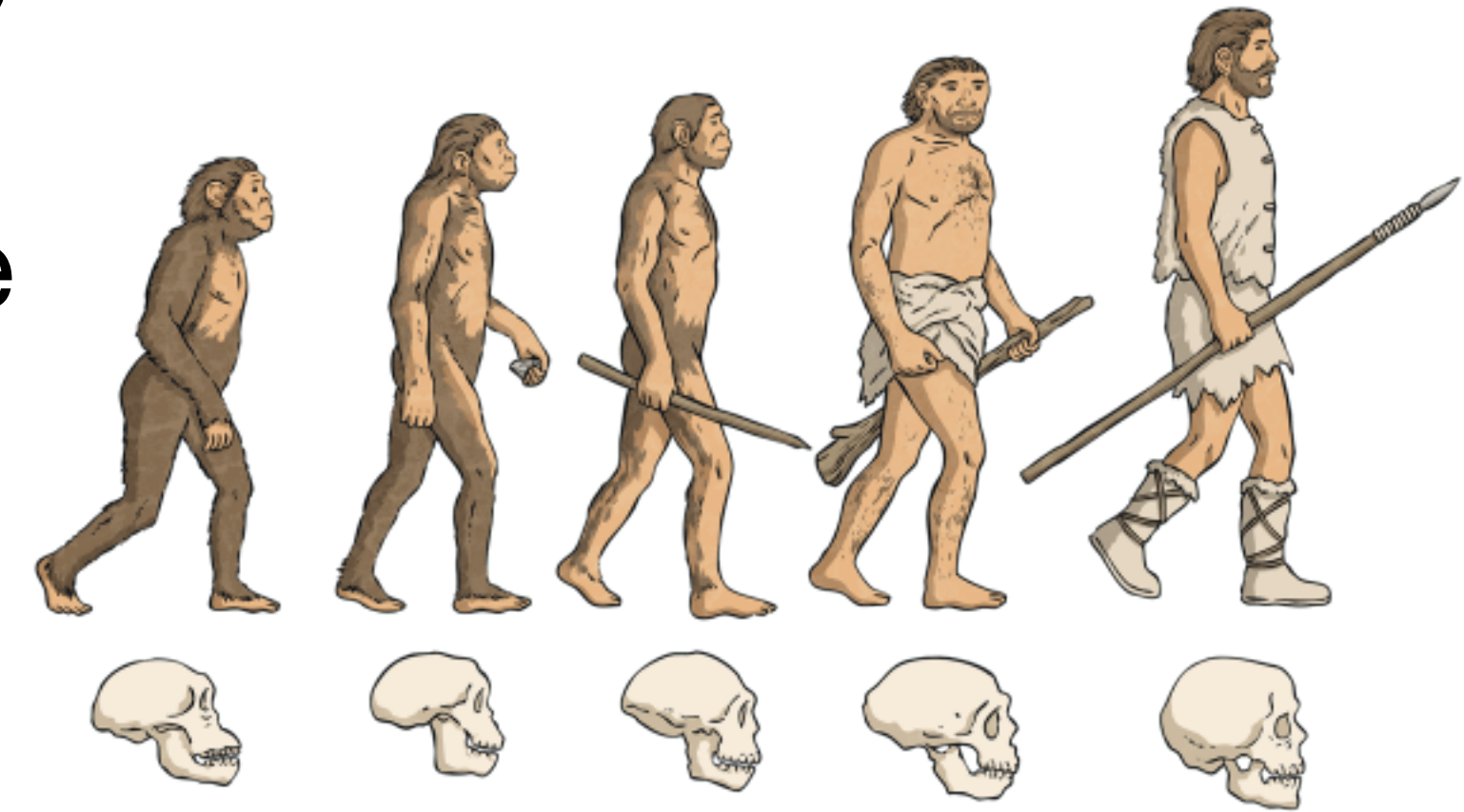


Breakthrough 4. new capabilities

- Model of mind : self and others
- Planning for future — anticipate/model future need
- Transfer of knowledge/skill
 - training by learning intent as well as motions
 - imitation learning

Breakthrough 5: (humans) labels and language

- Geological event split dry grassy plains from lush protective forests
 - walking upright -- needs bigger brain
 - managing fire, cooking -- more energy from less food
 - exposure -- need to cooperate
- Human brains are bigger, denser, more connected than earlier primates
- Two new regions of the NC for processing language
 - Broca's NC area controls production: speaking, writing, sign language
 - Wernickes NC region controls understanding: written, spoken, signed ...



Breakthrough 5: advantage of language

- Humans born with learning ability and curriculum for learning language
 - language is not hard wired
- Language enabled
 - accumulation of knowledge and skills across generations
 - sophisticated tools
 - more inventions

Summary

- Natural intelligence evolved as survival of the fittest.
- Capabilities incrementally developed bottom up shaped by environmental pressures.
- Sources of learning increased in complexity
 - from own actions -- trial and error (bilaterals) >>
 - from imagined actions -- trial and error (vertebrates) >>
 - from others actions -- imitation learning (early primates) >>
 - from others imagined actions -- (early humans) Language was last to appear in natural intelligence
- Language is the last to appear in natural systems
- Language is the first to appear in logical systems

Can the analysis of break throughs answer the questions motivating the FOL project ?

- How can minds forget? (synaptic strength fades, chemicals disperse)
- How we can be interrupted? (multiple concurrent activities, mechanism for focusing attention / moving control between simulation and action)
- How does perceptual information get to be cognitive information? (sensory nerves >> neural patterns, change of synaptic strength)
- How we can comprehend infinite objects such as the set of numbers? (symbols and language, (meta) simulation?)
- How we can reason about other's knowledge, motivations? (mental models)
- How can we “know” in an ever changing environment? (model revision, exchange of knowledge, episodic memory, time)

The FOL System

Towards MRS

FOL project vision

To build (minds of) autonomous agents that can come to “know” things and interact with their environment

- Such an agent should
 - reason -- not just explain reasoning
 - refer -- not just be a theory of reference
 - acquire new behavior without ‘brain surgery/reprogramming’
- The idea is to use logic like builders use mechanical or electrical engineering.
 - The theory tells us how to put what blocks together.
 - The physics makes it work.

FOL elements

- FOL Contexts
 - the logical core / building blocks — finite data structures
- FOL Systems
 - collections of contexts constrained by a Physics
- Inference rules
 - maps on FOL contexts or systems, moving through time
- The FOL engine
 - execution of rules, semantic simplification
 - Eval — mixed syntactic and semantic inference

FOL Contexts

- An FOL context is a triple: $C = \langle L, SS, F \rangle$
- L is the language of C :
 - $\langle \text{Sorts, Funs, Rels, Consts} \rangle$
- SS is the simulation structure of C over domain D
 - $\langle \text{Reps, SAtts, FAtts, RAtts, CAtts} \rangle$
 - SS determines maps: $\text{Exp}(L) \rightarrow D$
 - SS determines a “partial model”
- F is the set of facts of C :
 - $\{ \dots \langle \text{fname, wff, justification} \rangle \dots \}$

	$\leftarrow$	L	
FOL		-----	F
	$\leftarrow$	SS	

Wristwatch example

- A wristwatch context -- impressionistically

$$\text{wwC}[n] = \langle \underbrace{\langle \text{now} \rangle}_{[n]} \{ \} \rangle$$

- L_{wwC} -- the constant now
- SS_{wwC} attaches **now** to a datastructure holding a natural number value
- There are no facts

FOL Inference Rules

- An FOL inference rule is a map on contexts that preserves consistency
- $C \text{ -ir-} > C' = \text{fir}(C)$
- Example: $\text{wwC}[n] \text{ - tick -} > \text{wwC}[n+1]$
- Examples of rule classes
 - Adding a new symbol
 - Attaching a semantic object to a symbol
 - Forgetting a fact, an attachment, a symbol
 - Adding a computational consequence to the facts
 - Proof rules

FOL Engine — support for mechanizing

- Simplify — syntactic evaluation, uses inference rules, facts/equations until none apply.
- SVal — semantic valuation
 - $C \text{ -SVal}(e,tc) \rightarrow \text{SVal}(C,e,tc)$
 - adds a fact equating expression, e and its semantic simplification, tc
 - uses semantic attachment to move between syntax and semantics
- Eval — combines Simplify and SVal

FOL Physics

- An FOL system, Sys , is a finite collection of labeled contexts
- An FOL system is consistent if its each of its component contexts is consistent.
- An FOL physics has
 - a collection of FOL systems — Sys in SysS
 - a predicate P
 - a collection of inference rules, ir , such that
 - $P(\text{Sys})$ implies $P(f_{\text{ir}}(\text{Sys}))$
 - for Sys in SysS and $f_{\text{ir}} : \text{SysS} \rightarrow \text{SysS}$

Theory of Time Physics

- Theory of time contexts

$$\text{totC}[m, l] = \underset{\substack{| \\ [m]}}{<} \underset{\substack{| \\ [l]}}{< \text{now then } l e} \underset{\substack{| \\ <=}}{> \{ l e(\text{then}, \text{now}) \} >}$$

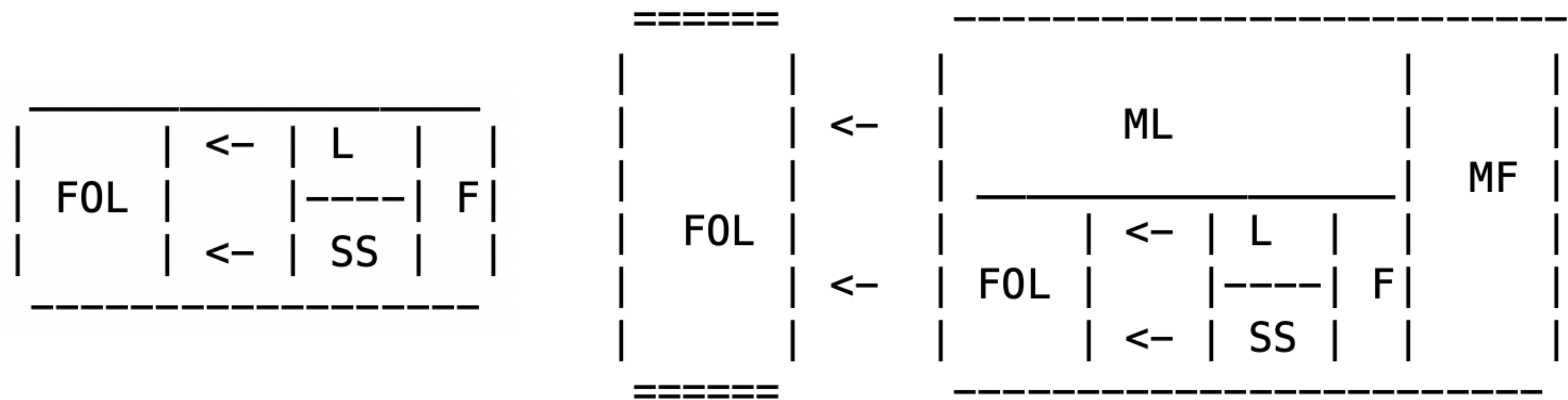
- A theory of time physics:
 - $\text{totP}[n, m, l] = \{ \text{wwC}[n], \text{totC}[m, l] \mid \text{s.t. } l, m \leq n \}$
- update-then rule: $\text{totC}[m, l] - \mathbf{update-then} \rightarrow \text{totC}[m, m]$
- **tick** and **update-then** lift to totS
- look rule: $\text{totS}[n, m, l] - \mathbf{look} \rightarrow \text{totS}[n, n, l]$
- Theorem: Any combination of **tick**, **update-then**, **look** preserves totS

Meta Systems, Meta Physics

- $\text{Meta}(\{M, C\})$ means that M is a meta theory of C .
- Exist universal meta context M such that $\text{Meta}(\{M, C\})$ for any context C , including M
- If M is a universal meta context then lifting object level deduction and semantic valuation at the meta level preserves Meta.
- If M contains a fact such as "the number of facts is 3" then adding to the facts of C is not Meta preserving.
- Reflection principles move information between levels

Self reflection

- The FOL Meta physics shows how a logical system can look at itself — - attach the constant META to M!
- The FOL engine is itself the natural simulation structure for the universal meta theory .



What intelligence features can an FOL system exhibit?

A few examples

- Basic sense ; evaluate ; act — using semantic attachment, eval
 - Example: A camera context with symbol **camera** attached to an actual camera. Rule: Activate switch if red sensed.
- forgetting — a consistency preserving rule on contexts
- simulation of the world to predict result of actions — using semantic evaluation
- simulate self to understand intent - universal meta context
- label / language—This is the easy part: For Contexts — new symbols can be learned (being told ...) and their properties expressed.

What is missing

- A real notion of time -- rhythm, when to act
 - Natural systems have builtin notions of time ticker cells, periodic (chemical/electrial) processes
 - How can a mechanized reasoning system have an internal clock?
 - What is time? How can I tell if a system has a notion of time?
- Various affects seem important for natural intelligence
 - motivation (intent,desire,fear), focus of attention
 - in natural systems the underlying mechanism seems to be chemical (hormones)
- What about imagination?

Some questions and conclusions

Open question: characterizing the intelligence of MRS

- Are there interesting hierarchies of capabilities.
- The different breakthrough stages of brain / intelligence development have different capabilities.
- Is there a useful corresponding notion for MRS?
- Semantic attachment + reflection is very powerful.
 - What are meaningful restrictions?
 - on computational power
 - on sensing/action capability
 - Is there a form of semantic attachment without language?
 - Is language necessary for reflection?

Concluding

- Logic explains/validates reasoning,
 - based on rules for reasoning
 - a closed system
- Intelligence involves
 - data processing / reasoning (a brain)
 - sensing, acting, feedback, mental experiments,
 - a notion of time
- The proposed idea of mechanized reasoning systems is intended to help
 - understand natural intelligence
 - build active entities that exhibit desired intelligence

The end!

Thanks

References

[Brooks] Rodney Brooks, *AI: great expectations. Manufacturing Engineering*. 1988.

[Bennet] Max S. Bennet. *A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs That Made Our Brains*. Mariner Books, October 24, 2023

[Weyhrauch] Richard W. Weyrauch. Prolegomena to a theory of mechanized formal reasoning. *Artificial Intelligence*, 13:133-170, 1980.