## INTRODUCTION

Cog sci = interdisciplinary study of the mind, inherently very recursive, thinking about thinking, using modern methods to answer classic questions (where does knowledge come from, what is the nature of thought?, are there uniquely human aspects of cognition?)

* Philosophy, psychology, linguistic, computer science, neuroscience, anthropology

Rationalist vs. empiricist:

|  |  |
| --- | --- |
| RATIONALIST | EMPIRICIST |
| * Knowledge comes partly from **reason**, some aspects of which may be **innate** * Where does knowledge come from? **Some knowledge is innate, some knowledge has to be derived from reason independently of the senses** * What is nature of thought? **Language-like**, propositional, logical * Are there uniquely human aspects of cognition? **Yes** | * Knowledge comes from **experience** * Neither principles nor ideas are innate * Where does knowledge come from? **ALL from experience, sense-based experience, cant come from intuition and reasoning alone** * What is nature of thought? **Experience-like**, imagistic, associative * Are there uniquely human aspects of cognition? **Not fundamentally** |

|  |  |  |
| --- | --- | --- |
| Rationalist | Empiricist | (year) |
| Plato  Leibniz  Boole, Frege  Turing  Chomsky  Symbolic AI  Rational Models | Aristotle  Locke  Skinner  Neural networks  Embodiment | ~350 BC  ~1700  ~1860  1950  1959  1980s  Now |

CRUM = computational-representational understanding of mind

* How cog sci tries to portray/understand mind, AI
* Thinking about mind like a computer 🡪 AI, Turing Machine, Behaviorism (don’t believe in mental rep in mind)
* Cognition = information processing
  + Information processing can be formalized 🡪 operates on represenations (symbols that have meaning) --> cog sci can be represented separate from physiology/biology 🡪 can represent with symbols, computers instead

## INNATENESS

Aristotle = **THE BLANK SLATE**

* Empiricist
* The mind is a “blank slate”, empty potential, nothing is there yet

Plato = **LEARNING IS REMEMBERING**

* Rationalist
* There is innate knowledge: we know things that we could not have learned, things are learned NOT just from experience, so the soul must have understood them for all time
* “learning is remembering” (the play example, where the boy discusses the square and its length/area)
* Plato’s problem: how can humans know so much when contact with world is so brief personal and limited?
  + Poverty of the stimulus = limited experience (our contact with the world is brief, personal, and limited, couldn’t have learned that much from such little stimulus)
  + We **remember** facts from before our birth, knew it already 🡪 the only way we could come to know so much is if we knew it already
  + Argument from poverty of stimulus
    - We have some piece of knowledge, K 🡪 our experience is too limited to have learned K from experience 🡪 therefore we must have always known K
* Soution:
  + What exists in the world are SHADOWS of ideal forms, but they are similar/shadows enough that they remind us of the ideal forms in our memory/soul (the world triggers our memory 🡪 relearn things)

Locke = **BUILDING KNOWLEDGE FROM EXPERIENCE**

* empiricist
* Neither principles nor ideas are innate
  + There is another way men come to a universal agreement on certain truths
* “White paper” (like Aristotle)
  + No character, no ideas
* Experience is either: sensation and reflection
  + Ideas comes from sensation or reflection: external objects or internal operations of our minds
  + Sensation: Depend on our **senses** for an understanding of external objects and to then form ideas from it, distant perception of things (i.e. white, soft, hard, bitter..)
  + Reflection: mind **contemplating,** on its own operations within itself (i.e. thinking, perceiving, knowing, beliving…)
* People gradually get more experience, mind THINKS IN PROPORTION TO THE MATTER IT GETS FROM EXPERIENCE TO THINK ABOUT
  + More knowledge/observations 🡪 become more familiar with certain objects 🡪 advances our thinking, ideas, understanding, can distinguish, etc
  + Complex ideas come form from simple ideas (i.e. unicorn = horse + horn, even though we don’t experience it, we have the idea, can build these ideas up with more experience, advances our thinking)
* Empirical evidence

Leibniz = **VEINS IN THE MARBLE**

* (on locke) Experience is necessary(a little) but **not sufficient** to account for knowledge
  + Mind must have *something* to begin with to go beyond and understand things
* This separates humans from beasts/animals
* Necessary truths
  + Our predisposed knowledge allows us to know necessary truths, and that we know they hold universally
  + Knowledge that goes beyond all your expierences!!!!!
  + We can’t expierence everything, but still know certain thigns
* Seeds of eternity
  + Minds contain these seeds of eternity that lets us GO BEYOND the merely empirical, reveal something divine/eternal
  + They are flashes of light hidden inside of us
* Beasts are purely empirical
  + Guided solely by instances, don’t form necessary propositions (from an analysis of notions)
* Veins in marble
  + Marble is predisposed to hercules’ form more than any other form, those veins encourage hercules’ formation
  + **= the mind is predisposed to certain ideas and information, it just takes experience and reason to uncover and refine them**

|  |
| --- |
| Leibniz vs. Locke |
| Locke is arguing his based on empirical evidence (what you sense = sensation, your inner workings of your mind, etc. it’s white paper, and you gradually gain knowledge from experience which leads you to more complex ideas/thinking)  Leibniz is arguing that you’re already predisposed to certain things, you just need a trigger/stimulus to learn it. You know it has to be true for sure. While locke’s evidence could always be wrong because it is empirical |

Induction vs. Deduction

* INDUCTION
  + Generalizing **beyond the data** given
  + Guided by **bias**
  + Conclusions could be **wrong**
* DEDUCTION
  + Known truths 🡪 new truths
  + Conclusions are **certain**
  + Syllogism
    - Premises 🡪 conclusion (all a’s are b’s, all b’s are c’s, so all a’s are c’s)

## ALPHABET OF HUMAN THOUGHT

Need a theory of thought, that will PREDICT human thoughts and then EXPLAIN human thoughts! = A mathematical theory of thought should predict and explain how we reason

Aristotle: (**CATALOGING THE SYLLOGISMS)**

* Wanted to create a system for deductive reasoning
* His project was to create a catalogue of valid syllogisms (deductive reasoning)
* Validate syllogisms by creating all possible syllogisms, determining which are valid/invalid
  + Limited it to sentences, only considered sylls with 2 premises and 1 conclusion
  + He only argued for the FORM of the argument
* Have premises 🡪 reach a conclusion because premises are so
* All = universal, some = particular (either confirm or deny them)
* Achieved a systematic approach and general rules (like no valid syllogism has 2 negative premises)
* CONS: too restrictive, constricting, not looking at the bigger picture/starting too small
* His approach makes predictions in a CERTAIN WORLD, though we need to make conclusions in an uncertain world, need to take a more rational approach

Leibniz: **(THE VISION)**

* Dream/wonderful idea = **an ideal language that perfectly represents relationship between our thoughts (**because LANGUAGE is an imperfect mirror of human thought, we don’t actually say exactly what we think perfectly); we are perfect though, just need notation that does all of the work to aid the clarity of thought
* Fundamentally optimistic, world we live in is the best of all possible worlds! World is not accidental, not undetermined (necessary truths, true in all possible worlds, that is how we have these necessary truths because of the veins in the marble…etc)
  + He believes that people are mostly benevolent/cooperative 🡪 but we are hampered in cooperation b/c of language, it doesn’t perfectly represent our thoughts (it’s clumsy, not effective) so our reasoning is obscure
* To create the ideal perfect language that perfectly represents the relationship between our thoughts, we need:
  + A compendium of all knowledge
  + Describe all knowledge in terms of key underlying notions 🡪 provide symbols for those notions
  + Reduce rules of deduction to manipulation of these symbols
  + == AI!!!
* “universal characteristic” is a language that aids the clarity of thought 🡪 notation does all the work

Boole: **(AN ALGEBRA OF THOUGHT)**

* Propositional logic
  + Uses Classes or sets of objects
    - X = white, y = sheep, xy = white sheep
  + What is xx? Xx = x (foundational rule)
  + Nothing can both belong and fail to belong to a class x (Aristotle’s principle of contradiction)
  + Applying boole to Aristotle:
    - A = AB, B = BC 🡪 A = AB = A(BC) = (AB)C = AC 🡪 A = AC
  + Became a basic component of computer programing (x = 1 is true, x = 0 is false, xy = 1 means both are true, x(1-y) = 0 means if x is true then y is true)
  + 1 = all things, true; 0 = none/nothing, false
* Can’t explain all thoughts cleanly though, modern logic can express more complex thoughts
* Clear deduction
* “everybody loves somebody” doesn’t work (it’s at the same time, universal & particular in the same sentence)
  + Multiple generality “every” vs “some”, need to differentiate between the two, Boole doesn’t do that

### LOGIC

Propositional Logic

* + X ^ Y = X and Y
  + X v Y = X or Y
  + X 🡪 Y = if X then Y
  + ~X = not X
* Build complex formulas with Boole’s logic
* Create TRUTH TABLES (if something is T, and other is F, then what is conclusion?) 🡪 TRUTH OF A FORMULA DEPENDS ON THE TRUTH OF ITS PARTS
  + Have the possible worlds, and models that are actually true
  + Assignment of truth value to propositions (the possible worlds) resulting in a formula that *is* true means it is a “model” of that formula
* Propositional logic has propositions = sentences that are defined by x and y, are these propositions true and in what worlds?
* Diff from predicate logic because predicates are used to describe objects, how properties of objects are true in what worlds

Frege

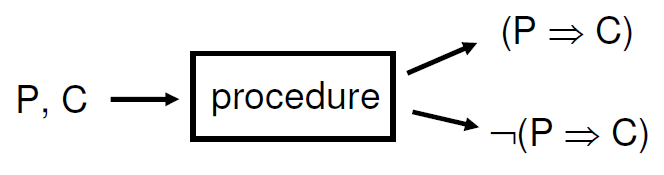
* Modern logic = begriffschrifft (a formal language aiming to capture relationship of thoughts) (“concept script”)
* FIRST ORDER PREDICATE LOGIC
  + Use predicates instead of propositions to denote properties of objects
  + P(a) = P is predicate and a is an object (i.e. Hairy(rex))
  + Variables are no longer propsitions, but just objects instead
  + Construct formulas same way as proprositional logic BUT it adds:
    - QUANTIFICATION: Two new symbols (quantifiers)
      * universal = A (“for all”) = all
      * existential = E (“for some” “there exists”) = at least one
      * they tell you how a predicate behaves for certain values (using quantifiers because of predicates, when do they occur and for what)
  + possible worlds = truth values assigned to all predicates applied to all objects
  + model is true dpeneding on assignment
  + LIMITED because you can’t say “for every PROPERTY P, there is some object that possesses that property” (we can say every OBJECT that is BLUE, can’t use quantifiers on predicates specifically)
  + We can talk about OBJECTS, can’t talk about PROPERTIES (all/some) 🡪 requires 2nd order logic
* Rules:
  + Modus ponens: always works, valid (if x then y is true, it is x, then it is y) (affirming antecedent) (if you know x, and you know if x then y is true, then conclude y)
  + Modus tollens: valid (if x then y, it is not y, then it is not x)
  + Affirming the consequent or denying the antecedent are INVALID
    - If x then y, it is y, it is x = FALSE
    - If x then y, it is not x, it is not y = FALSE
* Purely syntactic operations 🡪 LOGIC!!! Describe world with formulas
* SOLVES LEIBNIZ’S DREAMS b/c algebra of thought that yields valid inferences
* Disproved by Russel (russel’s letter to Frege)
  + Problem: set can be a member of itself
    - Extraordinary vs. ordinary example
  + Problem 2: not an efficient method, Frege’s method is NOT efficient (leibiniz wanted efficiency) (if you get it wrong, you can’t tell whether it’s a matter of laziness or you just did it incorrectly) (frege never provided a way to tell which is which, that the premises didn’t follow or that you’re just lazy)
    - If you fail to show that conclusion C follows from premise P, you don’t know if it actually follows or not. No procedure to tell you what follows what and how

### INFINITY

* Experience is finite, but we can conceive and portray the infinite
* Rationalist argument that challenges empiricism
* We go beyond the data given (Leibniz’s “seeds of eternity”)
* Infinity = large, important, positive concepts: GOD is infinite and all powerful, the ABSOLUTE
* CANTOR
  + Countably infinite, uncountably infinite (there’s no one size of infinity)
  + Uncoutnably infinite set of (real numbers, or anything) is larger than the set of countabley infinite set of (real numbers/anything)
  + Countably = set of all natural numbers, even numbers, odd numbers (natural #s consists of both even and odd set, but theyre all still the same size = infinite)
  + **the diagonal method**
    - cantor used it to show there is more than one size of infinity, later used by turing to show if leibniz’s dream (language of thought) was attainable
    - you can always create a package that is unlike all the others that it is formed from
    - take the complement, and the complement set is not in the actual set
    - generally more sets than there are integers 🡪 sets make it uncountably inifinite cause you can create infinite sets from a set
  + N is the set of all natural numbers, and all the possible subsets of N is UNCOUNTABLY INFINITE since there are infinite # of different subsets of N
    - Uncoutably infinite > countably infinite
* Platino and St. Augustine believed in Infinite = god; plato and aristotlte did not believe in the infinite
* Galileo, adding small parts to lengths, small gaps = infinite number

### THOUGHT AS COMPUTATION

* Still don’t have a method to determine whether a conclusion actually follows from a premise
* Hilbert’s *decision problem*
  + “Entscheidungsproblem”
  + Want a procedure that accepts premise P and conclusion C, and **determines whether C follows from P**



* + This would fulfill Leibniz’s dream because it would show the exact relationship between people’s thoughts and if something does work or not 🡪 didn’t work out b/c of turing proved that with diagonal method that you could not determine if conclusion C followed from premise C
* TURING
  + Used diagonal method to show that the procedure doesn’t exist
    - There are some claims that can’t be decided computationally
    - Using diagonalization to create a contradiction
    - Related to halting problem: *Given a program and an input to the program,   
      determine if the program will eventually stop when it is given that input.*
  + TURING MACHINE = formal model of computation (specific purpose)
    - Universal turing machine = model of general purpose computation (i.e. today’s computers, phones, etc) = can be programmed to simulate any other turing machine
    - Internal states are finite, tape is infinite (symbols read and written), finite set of rules that tell the machine what to do as a function based on what is on the tape
  + Turing machines could possible represent our way of computation because:
    - Some machines see 0, print 1: they are reflecting NOT just the input, so this is rationalist (can go beyond what you see, experience, so must have some internal computation or must have always had this information)
    - Some machiens see 0, print 0: reflect just the input, so empiricist (what you see, experience)

## THOUGHT COMPUTATION AND THE WORLD

* Symbol grounding problem
  + Symbols represent propositions of things in the world (abstract objects that stand in for real objects)
  + Need to know when facts about the world are TRUE, so symbols are also true still 🡪 how do symbols get their meanings? 🡪 so what is meaning itself?
  + Cognition is considered computation aka manipulation of symbols (but symbols aren’t concrete, they are semantic, how do we know that the symbols are actually connected to the things they refer to?)
  + N
* get our information through our senses about facts in the world which we are know are true 🡪 make new inferences from these facts about the world 🡪 inferneces make us take action to get more information
* so if our SENSES determine our facts about the world 🡪 we could potentially be mislead 🡪 **so what can we know with certainty?**
* **A major cog sci proposal is that cognition/thinking = computation 🡪 BUT THAT MEANS COMPUTATIONAL SYMBOLS MUST BE GROUNDED IN THE WORLD (symbols must actually be connected to the things they refer to in the world, which is meaning but what is meaning then?)**
* Descartes
  + Always looking for the truth
  + You can doubt the existence of your body, but can’t doubt the existence of your mind (body = from senses, mind = must exist because by thinking you establish its existence)
* Dualism: mind and body/world are different kinds of entitites ( made of different stuff)
  + Mind/matter cannot be reduced to one another
  + No explanation of minds in physical terms
  + Mental states cannot be reduced to neural states
  + Physical sciences cannot explain operation of the mind
* Materialism/Monism: mind and matter are the same thing
  + Mind can be reduced to matter, can be described in physical terms
  + Can explain our mind with sciences like physics & biology
* Mind-Body Problem:
  + If they are two different things, how does info from our senses get to our mind 🡪 and how does the mind cause the body to act?

## BEHAVIOR & THE MIND

* Boole
  + Deduction, *not induction*: truth is seen in what you see, no repetition of instances 🡪 the world is what you see, those are the facts laid out for you
* Wundt: father of experimental psychology, subjective introspection
  + Reaction time of pendulum: people reporting position of pendulum at click 🡪 perceived to be late
  + **Apperception:** process of making experience clear in consciousness 🡪 inferring unobserved mental process from observed behavioral data
  + **Brass instrument techonology**
  + Experimenters at this time relied on people’s reports of their experiences = so everything is highly **subjective 🡪 subjective introspection**
* Ebbinghaus: avoid introspection, **spacing effect, memory retention**
  + Recorded objective measures, not intropsection
  + List of items better remembered at Intervals > cramming
* **BEHAVIORISM:** Watson, behavior > mind, mind does not exist, black box, NO introspection
  + role of environment in explaining behavior
  + behavior > mental states
  + no substantial difference b/w humans & animals
  + purely objective
  + **EMPIRICIST** (knowledge from experience, knowledge is experience like, no human aspects of cognition)
  + Focusing on behavior allowed animal experimentation door to open
  + Need to just shape our envnt to understand our behavior, no need for mind/thinking theories
  + Study all this by animals if it’s the same, because learning principles in humans will appear in animals
* Classical conditioning: associating one stimulus with another for a response
  + Learning one cue (conditioned stimulus) is associated with another stimulus that has a natural reaction (unconditioned stimulus)
  + Associate stimuli paired in experience but not under your control
    - Ex. Pavlov’s dog & bell
      * Dog salivates for food (unconditioned stimulus), bring in conditioned stimulus (bell) whenever there is food 🡪 take away food 🡪 ring bell 🡪 salivate for bell
* Operant conditioning: action 🡪 reward/punishment
  + Learning that performing an action leads to a reward or a punishment, so more likely or less likely to perform an action or do it faster or slower
  + Associate actions with reward/punishment
  + Ex. Cats in puzzle boxes
    - Learned how to escape when there was a food reward
    - **Thorndike**
* Environment comes into play in behaviorism because it shapes people and explains their behavior (empiricist in nature 🡪 experience 🡪 behavior 🡪 knowledge and understanding of the world)
* Little Albert experiment
  + Classically Condition baby to fear mice, associate loud noise with it 🡪 teach it fear of mice
  + See if conditioning principles in non-humans apply to humans
* Skinner: radical behaviorist
  + No mental states at all
  + Skeptical that if we only believe the information from senses, what can we really know?
    - No conclusive evidence for minds, so how do we really know our knowledge is real
  + Skinner box
    - Operant conditioning
    - Mice, or other animals in a chamber
    - Can touch something (like a lever) that will elicit a response 🡪 positive reinforcement

## THE COGNITIVE REVOULTION

* Birth of cog sci 1956 MIT symposium; Before this, during all the behaviorism stuff, cog sci/mental states were rejected
* Tolman
  + Challenged behaviorism
  + Latent learning in rats: Rats in the maze
    - Some unrewarded, some rewarded every time, some not rewarded then rewarded after a while (3rd group learned more quickly)
    - Rats latently learned about the maze w/o reinforcement 🡪 but once reinforced, used that prior knowledge/learning to build on it
  + **Learning is driven more than reinforcement**
  + Pick up info, store it, use knowledge when relevant/useful
  + **Cognitive maps**
    - Mental representations (ex. In mice, faced in certain direction and had to get to the food)
  + Showed that even non-human learning has mental representations
* **MENTAL REPRESENTATIONS:** key feature to cog sci, distinguishes it from behaviorism
  + Behaviorism says: environment 🡪 behavior
  + Cog sci says: environment 🡪 representation/mental proessing 🡪 behavior
* Newell & simon
  + The Logic Theorist
  + First AI system: found proofs of mathematical facts 🡪 expressed in terms of inference rules 🡪 used heuristics
  + More elegant and efficient
  + Introduced idea that computers might be able to replicate human thought and problems, and human thougth can show us how computer programs should work
* George Miller
  + Magical number 7
  + Limit on capacity to process information; information processing constraints; constraint on human mental representation
  + Opens possibility of heirachically embedding chunks inside one another
* Lashley
  + Hierarchical structure of plans
  + Mental representations must have **structure**
* Chomsky
  + Hierarchical structure of language
* Computers will allow representing of all of this

## LANGUAGE

* Hierarchically structured 🡪 related to millers chunks, chunks of mental representations that are hierarchically structured in the mind
* Whitehead
  + Worked with Russell, built on Frege, LOGIC
  + WHITEHEAD VS SKINNER
    - Verbal behavior/language is special, an ability that differs from others
    - Mental state encoding absence of something else, so how can you know/say that still
    - Classical challenge to behaviorism (language is special and uniquely human)
* Language is INFINITE
  + Relies on mental representations (against behaviorism)
  + Uniquely human and innate basis (against empiricism)
  + Potentially infinite (against empiricism)
* Whitehead vs. skinner again: skinner published book trying to explain language in behaviorist terms about 20 years later

🡪 argued against by CHOMSKY

* CHOMSKY
  + Rationalist
  + Language reflects knowledge and can only be referenced in terms of mental representations
  + Some of this knowledge has to be innate (veins in the marble idea)
  + Language is not just a stream/string of words 🡪 HIERACHICALLY STRUCTURED
* Know SOUND, SYNTAX, AND MEANING of language
  + We have a large amount of linguistic knowledge w/o knowing it, unaware
* 🡪 so we know the general rules of knowledge 🡪 can apply these rules to new things 🡪 infinitely generative
* So how do we learn language? Language acquisition problem (plato’s problem of how can we know so much from so little?)
  + Language is hierarchically structured
  + That structure is NOT learned through from our linguistic input as children 🡪 poverty of stimulus
  + So that knowledge of structure must have been innate, some language knowledge must be innate
  + We learn language with negative evidence about what is WRONG, and the rest we must have some innate knowledge about what is possible
  + through what structure
* LANGUAGE ACQUISTION DEVICE
  + Universal grammar
    - No negative feedback for some input, how do we know its right? Or what structures are possible? Or what can be true models in this world?
    - Innate knowledge
    - Patterns that are found in all/many languages
      * Sound, color

## THE DISCIPLINE MATURES

* Symbolic vs. imagistic representations
  + Symbolic
    - RATIONALIST VIEW
    - Language-like, propositional
    - Frege
    - Chomsky
    - Newell & Simon
    - **ELIZA & SHRDLU**
      * **Eliza**
        + Wizenbaum
        + Conversation-simulating chatterbot, doesn’t actually understand though
        + Looks for keyword/pattern in whatever person says, then applies rule to whatever that comment was, then returns that transfomartion (if keyword not found, just repeats something generic)
        + Mimicked understanding
      * **SHRDLU**
        + Terry Winograd
        + Modeling natural language understanding
        + Language linked to action: seeing language as a way of activating procedures within the hearer
        + Occurred in micro-world: artificially simple situations
        + Human types something into the computer, giving some command, and **SHRDLU** interprets what the human said. Answers the question. Determines what “it” is, what block they are talking about. Does SHRDLU understand what a RED BLOCK is though? It knows that it is taller, or its shape, and where to put it. Knows things about its world, and builds upon them, also had some sort of memory base.
        + 3 components/procedures

Syntactic analysis

Semantic analysis

Perception and inference: consult world for answers

* + - * + uses language to report on envnt and to plan action
        + illustrates how abstract grammatical rules might be represented in cog system, and how it is integrated with other info from envnt
        + **SHRDLU designs illustrates strategy in cog sci (breaking things down into compoentns, each with some information processing task)**
        + Information-processeing tasks are all implemented algorithmically (used to explain all of SHRDLU’s different procedures)
        + **Links symbolic understanding to action in the world**
      * These experiments are saying mental representations take a language-like symbolic form
  + IMAGERY/IMAGISTIC
    - These experiments that mental representations could be imagistic instead
    - Empiricist view
    - How do we understand information and information processing?
    - Imagery debate = information processing underlying conscious experiences
    - SOME info processing must involve operations on geometrically encoded representations
    - Debate is on whether different effects revealed by experiments (propeller, rotating image) on mental imagery CAN or CANNOT be explained in terms of digital information processing models
    - Experiments:
      * Airplane, propeller example (Kosslyn); scanning
        + Scan across a mental image, latency of scan is measured (measure of time delay experienced in a system)
        + Map scanning experiment
        + Linear relationship b/w response latency and distance scanned on image
        + 🡪 properties of mental representations and spatial extent/constraints
      * Mental representation of rotation (imagistic mental representations)
        + Shepard & Metzler experiment (rotating figures, are they the same or not?)
* David Marr
  + Introduced idea of different **levels of analysis for information processing systems**
  + At physical level, answer questions using tools of physical sciences
  + At abstract level, answer question using computational terms
  + Incorporates both imagistic and symbolic, it’s a continuum
  + = THUS MIND CAN BE UNDERSTOOD VIA THE DIFFERENT LEVELS (relate this to mind body, dualism, mind and matter, turing software & hardware)
  + Still shows you hierarchical structure of perception
* **Marr’s 3-level framework**
  + 1. Computational level
    - “goal of information processing” “what system is up to”
    - Translate general description of cog system into specific account of particular information-processing problem that is configured to solve
    - Identify the constraints that hold upon any solution to that information processing task
    - = job of individual cog systems is to transform one kind of info (like info coming in through sensory) into another type of info (like info about what objects in evnt)
    - = computational analysis identifies info with which cog system has to begin (input) and info with which it needs to end (output)
    - Ex. Functional goal of the visual system is to determine the shape of objects in the world; flying
  + 2. Algorithmic level
    - “software”
    - How we process information
    - How cog system actually solves the specific information-processing task identified at computational level
    - How input 🡪 output
    - Algorithms that effect transformation
    - Ex. 3d representation; curved wings & aerodynamics
  + 3. Implementational level
    - Find a physical realization for algorithm
    - “hardware”
    - Identify physical structures that will realize the representational states over which the algorithm is defined
    - Ex. Neurons; feathers
  + Systematic approach to tackling how to combine and integrate different levels of explanation
  + How different levels of explanation connect up with each other
  + Three different levels for analyzing cognitive systems
  + Cognition is to be understood in terms of INFORMATION PROCESSING
  + TOP DOWN analysis

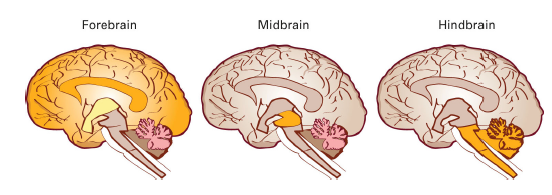
|  |  |  |
| --- | --- | --- |
| COMPUTATIONAL THEORY | REPRESENTATION AND ALGORITHM | HARDWARE IMPLEMENTATION |
| * Goal of computation? * Why appropriate? * What is the logic of the strategy by which it can be carried out? * What is the model trying to accomplish?   **Characterize the problem as a particular type of computation (what is being computed, what does it do, and why?)** | * How can computational theory be implemented? * What is representation of input and output? * What is algorithm of transformation? * What sort of processes are needed?   **Choice of algorithms for implementing the computation 🡪 internal representation 🡪 abstract formulation of how computation is carried out**  Ex. psychophysics | * How can representation and algorithm be realized physically? * What mechanism is needed to implement algorithm?   **Algorithm is implemented and physically realized 🡪 concrete formulation of how computation is carried out**  Ex. Neuroanatomy, neuro at the physical level is always implementational |

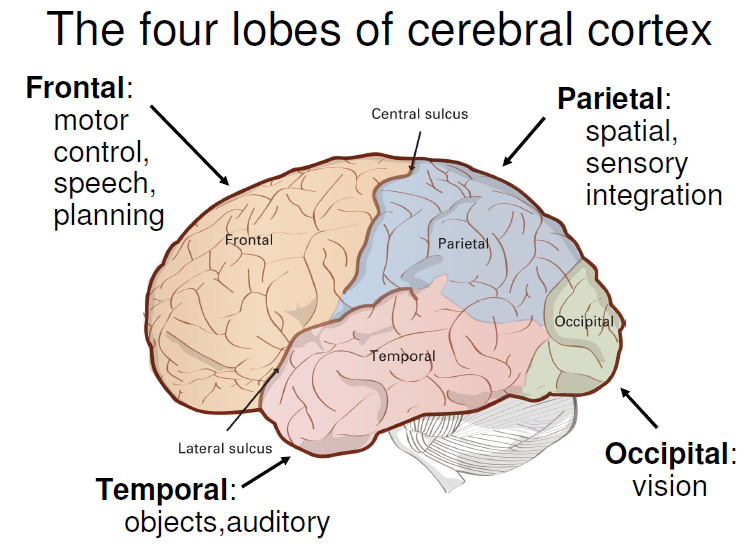
* **Marr’s model of vision**
  + Based it off of Elizabeth Warrington’s work on patients with damage to parietal cortex (lesions = problems with perceptual recognition)
    - Marr concluded that info about shape of object is processed separately from info about what the object is for and what it is called
    - Concluded also that visual system can deliver specification of shape of an object even if object is it not recognized
  + Five different stages: grey level, raw primal, primal sketch, 2-d, 3-d
  + AT COMPTUTATIONAL LEVEL: Basic task of visual system is to derive rep. of 3-d shape and spatial arrangement of object in a form that will allow shape to be recognized
    - Rep of object shape should be on an object-centered rather than an egocentric frame of reference
  + ALGORITHMIC LEVEL: how exactly is the input and output information encoded?
    - How the objects reflect light 🡪 primal sketch, areas of darkness brightness make some retinal image
  + IMPLEMENTATION LEVEL: produce representation that visual system must produce, depending on vantage point
  + Developing each stage, getting each component, to understand entire feature

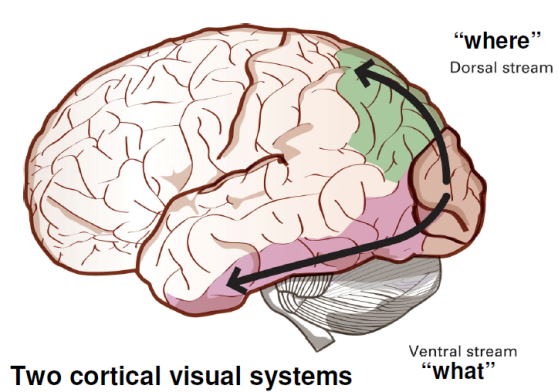
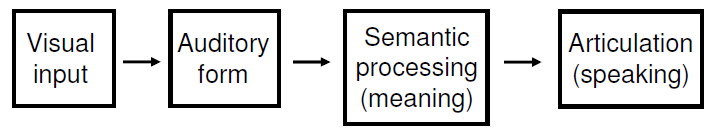
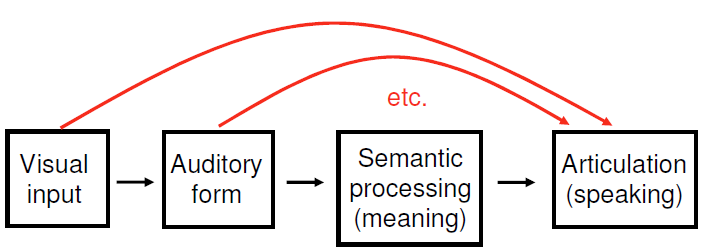
## THE TURING TESTS AND ITS CRITICS

* Theories of the mind often driven by current technology
  + Digesting duck
  + Mechanical man/turk
  + Water technology
* **The physical symbol system hypothesis** = a general approach to thinking about the mind
  + **Newell & Simon** 🡪 given award for fundamental contributions to computer science: created Logic Theory Machine and General Problem Solver
    - Programs that developed general strategies for solving formalized symbolic problems
    - **Delivered a manifesto for a general approach to thinking about intelligent information processing** 🡪 applied to study of human mind and emerging field of artificial intelligence = PHYSICAL SYMBOL SYSTEM HYPOTHESIS
  + All sciences are governed by basic principles (laws of qualitative structure) (i.e. biology 🡪 cells are basic building blocks of all living organisms)
    - Physical symbol system hypothesis = law of qualitative structure for study of intelligence (basic principles for the study of intelligence/mind)
  + **A physical symbol system has the necessary and sufficient means for general intelligent action**
    - takes physical patterns (symbols), combining them into structures (expressions) and manipulating them (using processes) to produce new expressions.
    - **Core idea:** problem-solving/thinking should be understood as the rule-governed transformation of symbol structures
    - **Human thinking = kind of symbol manipulation**
    - **Machines = can be intelligent**
    - 1st claim: nothing can be capable of intelligent action unless it is a physical symbol system (since humans are capable of IA then the mind MUST be a physical symbol system) 🡪 physical symbol system hypothesis is a constraint upon any possible mental architecture
    - 2nd claim: there is no obstacle to constructing an artificial mind, provided one tackles the problem by constructing a physical symbol system
  + Symbols = physical patterns 🡪 combine to form complex symbol structures 🡪 can be manipulated by physical symbol system 🡪 generating and transforming complex symbol structures are processes that are also symbols/symbol structures
  + 1. Symbols are physical patterns
    - For each symbol 🡪 there is a physical object
  + 2. Symbols combined to form complex symbol structures
    - Combining is governed by rules
  + 3. Physical symbol system contains processes for manipulating symbols and symbol structure
    - THINKING is the transformation of symbol structures according to rules
    - We solve problems by transforming symbol structures
    - Newell & simon claim that INTELLIGENCE & INTELLIGENT THINKING is the ability to solve problems 🡪 the ability to work out, which option best matches certain requirements/constraints
      * Search space: number of options and choices, which one is the best? Search-spaces are represented in terms of states (initial states and then a set of permissible transformations of that start state 🡪 governed by certain rules) 🡪 find a solution state/solve the problem
      * means-end analysis (GPS = general problem solver program)
        + intended to oconverge on solution state by reducing diff b/w current state and goal state: first evaluate diff b/w goal & current 🡪 identify transformation that reduce that diff 🡪 check that transformation can be applied to curr state 🡪 apply and do it again
        + = HEURISTIC SEARCH (make space search process shorter/quicker)
* The Turing Test
  + **Determine if physical symbol systems/machines actually understand using TURING TEST (since we’ve determined that physical symbol systems and machines are capable of general intelligent action/thinking/problem solving)**
  + Test of a machine’s ability to exhibit intelligent behavior, equivalent or indistinguishable from, that of an actual human
  + **Criterion: can convince you it’s a human**
  + Attempts to pass Turing test: Eliza, SHRDLU, Watson
  + Turing’s question: do machine’s think? How closely can a machine resemble typical human answers (doesn’t necessarily have to be correct)? Are there computers that can do well in the *imitation game*?
    - Imitation game: determine which of the 3 players is a man/woman (player C is the interrogator) (A tries to trick C, B tries to help C), can only communicate with written notes 🡪 Turing proposes replacing A with a computer as a woman and trick C 🡪 if computer can trick C, then computer is intelligent
  + Arguing on the basis of information processing systems = operating on formal symbols
* The Chinese room argument
  + **Argument against the Turing Test, says Turing Test cannot be used to determine if a machine can think, b/c it is inadequate criterion**
  + John Searle
  + Room full of Chinese symbols, given an English book with how to manipulate/respond these Chinese and what to write down 🡪 simulate a conversation. But that doesn’t mean he understands it?
  + Process whatever is given according to the program’s instructions and give some (correct) output
  + In either case, whether it is a person following the instructions or the computer running the program, it is the same thing. Both the computer and he are simply following a program, step by step, to reach some given output, **manipulating symbols by some process** 🡪 **simulating intelligent behavior**
  + HE doesn’t understand Chinese though, and thus we infer the computer doesn’t either
  + Searle may have BECOME the system as a whole, but still doesn’t understand Chinese
  + **Cannot give a computer a MIND, UNDERSTANDING, or CONSCIOUSNESS, regardless of how intelligently it may behave**
    - **Without understanding 🡪 can’t describe machine as thinking**

## THE TURN TO THE BRAIN

* Dualism
  + Descartes
    - Mind body problem, if mind/body are different things, how does info from senses get to mind and how does mind cause body to act?
    - Solution = Pineal gland = small endocrine gland (hormonal function)
* Phrenology (Franz Joseph Gall) = measurement of human skull, brain is organ of mind, and brain is localized = certain areas have certain function 🡪 based on the idea that human conduct could be best understood in neurological terms (by studying areas of the brain and determing emotions/character/thoughts) 🡪 steps toward NEUROPYSCHOLOGY
* Cesare Lombroso = criminality is inherited, not a trait of human nature
* Brain organization: lobes
* 



* Corpus callosum connects hempisheres
* Diencaphalon: Thalamus, pituitary, pineal
* Midbrain: primitive sensory/motor
* Hindbrain: balance, motor control, posture
* Central sulcus: divides frontal and parietal
* Lateral sulcus: divides frontal and temporal
* **Phineas Gage**
  + Railroad accident, iron went through his head, entered through side of his face, out of his head, STILL ALIVE
  + Intellectual facilities destroyed, not good with planning anymore (makes plans and disregards them immediately)
  + Frontal lobe = planning, executive function
  + Accidents can yield knowledge about functions of parts of brains, and functional organization of the brain, but that knowledge is **unsystematic and uncertain**
* **Hemispheres**
  + Language = left hemisphere
  + Cerebral asymmetries
    - Ex. Language: right handed people have more area in left hem associated with language
  + Contralateral organization
    - Info from left visual field processed in right hem, vice versa
    - What happens to visual info once it reaches visual cortex (occipital lobe)?
* **Two cortical visual systems: what(ventral, objects, temporal) vs. where(dorsal, space recognition, parietal) pathways**
  + **Ungerleider & Mishkin** = Two cortical visual systems
    - Did experiments on monkeys (cross-lesion disconnection experiments) = designed to trace connections b/w cortical areas and to uncover pathways along which info flows
    - Would take certain areas out, but communicate still worked through CORPUS CALLOSUM
    - Parietal cortex receives info from same visual cortex in hem, but opposite visual field, screwed up parietal cortex disrupts special recognition in opposite visual field (damage to right, disrupted left field)
    - Found in exp, that removing entire primary visual cortex did most damage, regardless of parietal cortex in given hemisphere
    - **Important in mapping out connectivity in brain 🡪 no single pathway for processing information**
  + Info about locating objects in space - “where” Dorsal stream to parietal; info about recognizing and identifying objects - “what” Ventral stream to temp
  + Parietal cortex = spatial cognition, receives most info from visual cortex in the same hem.
  + **Right** parietal cortex is responsible for spatial processing of info from **right** visual cortex 🡪 **right** visual cortex receives info from **left** visual field 🡪 damage to **right** parietal cortex disrupts special organization of **left** visual field
* **Hippocampus**
  + Spatial cognition
  + “place cells” theory
  + Neural correlate of cognitive maps
* **Words in the brain: serial vs. parallel models**
  + WORDS = different domain
  + IMAGING = different technique
  + 🡪 trace neural connections through successive interventions that build on each other
  + PET (positron emission tomography) = produces image of blood flow in body 🡪 correlates with functional processes
  + Use PET to determine word processing
  + Use subtraction method to subtract image of one task from that of another, to zoom in one function of ONE BOX
  + **SERIAL model**: neurological
    - Lexical info travels through fixed series (like auditory, then appearance, then meaning..etc)
    - visual input 🡪 auditory form 🡪 semantic processing (meaning) 🡪 articulation (speaking)
    - 
  + **PARALLEL model**: cognitive model
    - Different types of lexical info processing at once 🡪 several channels feeding into semantic processing
    - **The experiments they ran, by patterns of activation supported parallel rather than serial model of single-word processing**
    - 
* Accidents, surgery, imaging as sources of knowledge
  + Accidents can tell us what parts of brain have what function, how it relates, though random and uncertain
  + We can get knowledge through PRECISE SURGICAL INTERVENTION, in the form of targeted removal of specific brain areas to uncover connections between them
  + Can’t be certain for sure, it’s just a correlation
  + Can’t be confident though, it could just support the idea
  + Might be part of a bigger network, something may have affected something else, might be a more global thing, prerequisite to something else
  + Ex. Maybe what happened to Phineas gage was that his attention was shot, and then couldn’t plan (etc… you’re not sure)
  + Subtraction method: subtract the image of one task from that of another, to zoom in on a functional box
  + Damage to parietal = problems locating objects, damage to temporal = problems identifying/recognizing objects

## NEURAL NETWORKS AND CONNECTIONISM

* Neural computation, connectionism, as alternate view of computation in cognition
* Distributed representations, soft constraints
* McCulloch-Pitts model
  + They discovered neural networks
  + Use components that have some characteristic of real neurons,
  + Have binary states, inhibitory and excitatory
  + Outputs/inputs
  + Threshold
* Linear separation
* Perceptrons vs. multi-layer networks
  + Single-layer networks = perceptrons
    - Crucial limitations in what they can learn
    - Rosenblatt, studied single-layer networks = called them perceptrons
    - Cannot apply to functions that are NOT linearly separable
  + Multilayer networks
    - Receive inputs indirectly, as opposed to single layer networks which receive DIRECT input. Multi get input as output “hidden units”
    - Can compute any computable function, including non linearly separable
    - Perceptron convergence can’t be applied, so how to train/learn multilayer??
      * Paul Werbos
        + Backpropagation algorithm:

Error is propagated backwards through network form output units to hiden units 🡪 modify error in hidden unit 🡪 modify weights 🡪 propaged back down until input layer reached

Error is backwards, but activation is forward

* + - * + Organized into diff layers, not connected, have hidden layers 🡪 information enters via input 🡪 activates whatever units its connected to in next layer 🡪 etc
* Supervised vs. unsupervised learning
  + Supervised = network told what errors it is making
  + Unsupervised = network does not receive feedback
    - Hebbian learning b/c association b/w neurons can be strengthened w/o any feedback
    - Neurons that fire together, wire together
* Learning rules
  + Hebbian learning (Donald Hebb), how learning might take place int eh brain 🡪 learning is at bottom an associate process 🡪 increase in efficiency by association
    - Unsupervised learning
  + Perceptron Convergence Rule
    - Rosenblatt
    - Looking for learning rule that allows network w/ random weights & threshold to settle on a configuration of wegihts/threshold that would allow it to solve any given problem
    - SUPERVISED learning 🡪 whenever network produces wrong output for given input 🡪 adjusts weights and or threshold (process of learning is process of changing weights in response to error) 🡪 learning is successful when change given desired output
    - REQUIRES FEEDBACK ABOUT CORRECT SOLUTION TO THE PROBLEM THE NETWORK IS TRYING TO SOLVE
* Cognition as satisfaction of soft constraints